# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08

(11) International Publication Number:

WO 98/07864

(43) International Publication Date:

26 February 1998 (26.02.98)

(21) Internati nal Application Number:

(22) International Filing Date:

22 August 1997 (22.08.97)

(30) Priority Data:

23 August 1996 (23.08.96) 13 December 1996 (13.12.96)

**GB** GB

**A1** 

PCT/GB97/02273

9617671.4 9625996.5

(60) Parent Application or Grant (63) Related by Continuation

US Filed on

08/782,893 (CIP) 27 December 1996 (27.12.96)

(71) Applicants (for all designated States except US): MICRO-BIOLOGICAL RESEARCH AUTHORITY CAMR (CEN-TRE FOR APPLIED MICROBIOLOGY & RESEARCH) [GB/GB]; Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED [GB/GB]; 14 Kensington Square, London W8 5HH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHONE, Clifford, Charles [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). QUINN.

Conrad, Padraig [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG

(74) Agent: SCHLICH, George, William; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

#### Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECOMBINANT TOXIN FRAGMENTS

#### (57) Abstract

A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.

### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	1.T	Lithuania	SK	Slovakia
ΑT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑÜ	Australia	GA	Gabon	1. <b>V</b>	Latvia	S7.	Swaziland
AZ.	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazi)	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	1S	Iceland	MW	Malawi	US	United States of Ameri
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Vict Nam
CG	Солдо	KE	Kenya	NL.	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzsian	NO	Norway	ZW	Zimhabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ.	New Zealand		Simonowe
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania	•	
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation	•	
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sti Lanka	· SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

### RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the  $H_c$  domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the  $H_N$  domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The  $H_N$  domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the  $H_{\rm C}$ , amino acid residues 449-871 for the  $H_{\rm N}$  and residues 1-448 for the LC. Digestion with trypsin effectively degrades th  $H_{\rm C}$  domain of the BoNT/A to generate a non-toxic fragment designated  $LH_{\rm N}$ ,

which is no longer able to bind to and enter neurons (Fig. 1). The  $LH_N$  fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolat d LC.

It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

- (A) clostridial neurotoxin light chain:
- -a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.
  - (B) clostridial neurotoxin heavy chain H<sub>N</sub> domain:
- -a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- -the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- -the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- -the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H<sub>c</sub> domain.
- -a portion of the heavy chain which is responsible for binding of the native

holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H<sub>N</sub> of a clostridial toxin heavy chain, whilst lacking the functional aspects of a cl stridial t xin H<sub>C</sub> domain.

For the purposes of the invention, the functional property or properties of the  $H_N$  of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a  $H_N$  domain or to the functions of a  $H_N$  domain are references to this property or properties. The second domain is not required to exhibit other properties of the  $H_N$  domain of a clostridial toxin heavy chain.

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH<sub>423</sub>/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polyp ptid of the inv ntion may be obtain d by expression of a r combinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain pr ferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and  $H_N$  components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the inv ntion described in an xample below, the toxin light

chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodim nt of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain  $H_N$  portion or a fragment or variant of a clostridial toxin heavy chain  $H_N$  portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the  $H_N$  domain. Teachings of regions within the  $H_N$  responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the  $H_N$  domain or fragment, though it too retains the function of the  $H_N$  domain. It is conveniently obtained by insertion, deletion and/or substitution of a  $H_N$  domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a  $H_N$  domain or fragment, (ii) a C-terminal extension to a  $H_N$  domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, link d for example by disulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is

significantly reduced in an in vitro assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated  $H_{\rm C}$  of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated  $H_{\rm C}$  of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated  $H_N$  of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and  $H_N$  sequences of botulinum toxin types A, B, C<sub>1</sub>, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then d termines the targ t for a polypeptide - immunoglobulin compl x. Alt rnatively, the

WO 98/07864 - 8 - PCT/GB97/02273

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall endop ptidas activity of the composition will, of course, also be d termined by the amount of the diviative that is present.

While it is known to treat naturally produced clostridial toxin to remove the  $H_{\rm C}$  domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described abov.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated LH<sub>423</sub>/A (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polyp ptide or a fusion prot in and comprises nucleotides encoding residu s 1-

1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated  $LH_{728}/B$  (SEQ ID NO: 20).

The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The  $LH_N/A$  derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the  $H_C$  domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recov r d from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide or if expression

hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH<sub>423</sub>/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H<sub>c</sub> domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant  $LH_N$  derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The  $LH_N$  enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a / to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this nom nclature,

LH<sub>423</sub>/A = SEQ ID NO: 2, containing the entire L-chain and 423 amino acids of the H-chain of botulinum neurotoxin type A;

<sub>2</sub>LH<sub>423</sub>/A = a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;

<sup>2</sup>L<sub>/2</sub>H<sub>423</sub>/A = a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain;

<sup>2</sup>L<sub>FXa/2</sub>H<sub>423</sub>/A = a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and

 $_2$ L<sub>FXa/2</sub>H<sub>423</sub>/A-IGF-1 = a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention d signated LH<sub>423</sub>/A;

- Fig. 3 is a graph comparing activity of native toxin, trypsin generated "native"  $LH_N/A$  and an embodiment of the invention designated  ${}_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) in an *in vitro* peptide cleavage assay;
- Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;
- Fig. 5 shows the transition region of an embodiment of the invention designated L/4H423/A illustrating insertion of four amino acids at the N-terminus of the H<sub>N</sub> sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H<sub>N</sub> sequence then begins ALN...;
- Fig. 6 shows the transition region of an embodiment of the invention designated L<sub>FXa/3</sub>H<sub>423</sub>/A illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated H<sub>N</sub> will be cysteine;
- Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated L<sub>FX=/3</sub>H<sub>423</sub>/A-IGF-1, a fusion protein; the IGF-1 sequence begins at position G<sub>882</sub>;
- Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated  $L_{FXa/3}H_{423}/A$ -CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position  $\Omega_{882}$ ;
- Fig.9 sh ws th C-terminal portion of the amino acid sequence of an

embodiment of the invention designated  $L_{FXa/3}H_{423}/A-ZZ$ , a fusion protein; th C-terminal ZZ sequence begins at position  $A_{890}$  immediately after a genenase recognition site (underlined);

show schematic representations of manipulations of

Figs. 10 & 11

polypeptides of the invention; Fig. 10 shows LH<sub>423</sub>/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig. 12

shows the trypsin sensitive activation region of a polypeptide of the invention;

Fig. 13

shows Western blot analysis of recombinant LH<sub>107</sub>/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH<sub>107</sub>/B; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH<sub>107</sub>/B.

The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO:

<u>Sequence</u>

1

DNA coding for LH<sub>423</sub>/A

2	LH <sub>423</sub> /A
3	DNA coding for 23LH423/A (Q2E,N26K,A27Y), of which an
	N-terminal portion is shown in Fig. 4.
4	<sub>23</sub> LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>26</sub> K,A <sub>27</sub> Y)
5	DNA coding for $_2LH_{423}/A$ ( $Q_2E,N_{26}K,A_{27}Y$ ), of which an N-
	terminal portion is shown in Fig.4
6	$_{2}LH_{423}/A (Q_{2}E,N_{26}K,A_{27}Y)$
7	DNA sadiau ( vi Balta vi sa
7 8	DNA coding for native BoNT/A according to Binz et al
9	native BoNT/A according to Binz et al
10	DNA coding for L <sub>/4</sub> H <sub>423</sub> /A L <sub>/4</sub> H <sub>423</sub> /A
11	DNA coding for $L_{Ex_a}/_3H_{473}/A$
12	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A
13	DNA coding for L <sub>EXa</sub> / <sub>3</sub> H <sub>473</sub> /A-IGF-1
14	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-IGF-1
15	DNA coding for L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-CtxA14
16	L <sub>FXa</sub> / <sub>3</sub> H <sub>423</sub> /A-CtxA14
17	DNA coding for L <sub>FXa/3</sub> H <sub>423</sub> /A-ZZ
18	L <sub>FX=/3</sub> H <sub>423</sub> /A-ZZ
19	DNA coding for LH <sub>728</sub> /B
20	LH <sub>728</sub> /B
21	DNA coding for LH <sub>417</sub> /B
22	LH <sub>417</sub> /B
23	DNA coding for LH <sub>107</sub> /B
24	LH <sub>107</sub> /B
25	DNA coding for LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>28</sub> K,A <sub>27</sub> Y)
26	LH <sub>423</sub> /A (Q <sub>2</sub> E,N <sub>28</sub> K,A <sub>27</sub> Y)
27	DNA c ding for LH <sub>417</sub> /B wherein th first 274 bases are

WO 98/07864

28

modified to have an *E. coli* codon bias

DNA coding for LH<sub>417</sub>/B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

#### Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain ( $H_c$ ) of botulinum neurotoxin type A. This recombinant product is designated the LH<sub>423</sub>/A fragment (SEQ ID NO: 2).

#### Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an E. coli codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique Kpnl restriction site. The remainder of the  $LH_{423}/A$  coding sequence was PCR amplified from total chromosomal DNA from  $Clostridium\ botulinum\$ and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polym rase amplified DNA (bas s 914-1138 and 1976-2616) and the r mainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gene was then fully sequenced and cloned into a variety of *E. coli* plasmid vectors for expression analysis.

## Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, E. coli harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth phase. Expression of the gene is then induced addition bγ isopropylthio-β-D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH423/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

The recombinant GST- LH<sub>423</sub>/A is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH<sub>423</sub>/A is recovered in the non-adsorbed material.

#### Construct variants

A variant of the molecule,  $LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of  $LH_{423}/A$  producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A.

Two further variants of the gene sequence that have been expressed and the corresponding products purified are  $_{23}LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a *Eco* 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the *Eco* 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, L<sub>14</sub>H<sub>423</sub>/A (SEQ ID NO: 10), which c ntains an additional four amino acids between amino acids 448 and 449 of LH<sub>423</sub>/A at a position equivalent to the amino terminus of the

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L<sub>FXa/3</sub>H<sub>423</sub>/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L<sub>/4</sub>H<sub>423</sub>/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of  $L_{FXa/3}H_{423}/A$  have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

Specific examples described are:

- (1)  $L_{FXa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2)  $L_{FXe/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3)  $L_{FXe/3}H_{423}/A$ -ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a genenase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing  $H_N$  function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide

incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the  $H_N$  domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely an lgGsT, and a C-terminal extension providing a ligand binding domain, namely an lgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

#### Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of  $_2LH_{423}/A$  ( $Q_2E,N_{26}K,A_{27}Y$ ) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant GST- $_2$ LH $_{423}$ /A ( $Q_2$ E, $N_{26}$ K, $A_{27}$ Y) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600  $\mu$ g/ml) is incubated at 37°C for 10-50 minutes with trypsin (10  $\mu$ g/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$ , 1.8 mM KH $_2$ PO $_4$ , pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

<sub>2</sub>LH<sub>423</sub>/A is more stable in the presenc of trypsin and more active in the in vitro

peptide cleavage assay than is  $_{23}LH_{423}/A$ . Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

#### Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH<sub>728</sub>/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polyp ptide,  $LH_{417}/B$  (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus

equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native  $\text{LH}_\text{N}/\text{A}$  .

A gene has also been assembled coding for a variant polypeptide,  $LH_{107}/B$  (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

#### **Construct Variants**

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Anoth r variant's quenc equival nt to nucleotides 691 to 1641 of SEQ ID NO: 21

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

#### Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-<sub>2</sub>LH<sub>423</sub>/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA11) reactive against a conformation dependent epitope on the native LH<sub>N</sub>/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na<sub>2</sub>HPO<sub>4</sub> 1.15 g/l, KH<sub>2</sub>PO<sub>4</sub> 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20  $\mu$ g of GST- $_2$ LH $_{423}$ /A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH $_{\rm N}$ /A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na $_2$ HPO $_4$  anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5  $\mu$ g (5X10<sup>-6</sup> g) to 5 picograms (5X10<sup>-12</sup> g). Aliquots of 0.5 ml were injected into mice intrap ritoneally and deaths recorded

over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti-  $GST_{-2}LH_{423}/A$  antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST-2LH<sub>423</sub>/A antiserum.

<b>Botulinum</b>	Toxin/mouse
	- UNITED WAS B

Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control
1	0	4	4	4	4	4	4
2	-	4	4	4	4	4	4
3	-	4	4	4	4	4	4
4	•	4	4	4	4	4	4 .

<u>TABLE 2.</u> Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.

#### **Botulinum Toxin/mouse**

Survivors On Day	0.5µg	0.0 <b>05</b> µg	0.0005µg	0.5ng	0.005ng	5pg	Control
. 1	0	0	0	0	o	2	4
2	-	•	•	-	-	0	4
3 .	-	-	-		•		4
4	•	-	•	•	•	•	4

#### Example 4

Expression of recombinant LH<sub>107</sub>/B in E. coli.

As an exemplification of the expression of a nucleic acid coding for a  $LH_N$  of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide  $LH_{107}/B$  (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in  $\it E.~coli~BL21~(DE3)$  (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for  $LH_{423}/A$ .

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

#### SEQUENCE LISTING

#### (1) GENERAL INFORMATION:

- (i) APPLICANT:
  - (A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
  - (A) NAME: THE SPEYWOOD LABORATORY LIMITED
  - (B) STREET: 14 Kensington Square
  - (C) CITY: London
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): W8 5HH
  - (A) NAME: FOSTER; Keith Alan
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
  - (A) . NAME: QUINN; Conrad Padraig
  - (B) STREET: Centre For Applied Microbiology And Research,
    Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E) COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
  - (A) NAME: SHONE; Clifford Charles
  - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
  - (C) CITY: Salisbury
  - (D) STATE: Wiltshire
  - (E), COUNTRY: UK
  - (F) POSTAL CODE (ZIP): SP4 0JG
- (ii) TITLE OF INVENTION: Recombinant Toxin Fragments
- (iii) NUMBER OF SEQUENCES: 28
- (iv) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (2) INFORMATION FOR SEQ ID NO: 1:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2616 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: single
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

								_		_	•					
ATG Met	GIII	Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC	AAC	TAT Tyr	Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn	GGT	48
GTT Val	GAC Asp	ATT	GCC Ala 20	IAL	ATC Ile	AAA Lys	ATT	CCA Pro 25	AAC Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	Pne	AAG Lys	ATT	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
GCA Ala 65	AAG Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
GAÇ Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT . Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720

		-		24	TT AF al As 45	,14 111	I AS	II AI	25	1 1y	r Gl	.u Me	t Se	er G 2	1y 55	Leu		768
, 0-			2	50	AG GA lu Gl	שם שב	u AI	26	5 PN	e Gly	y G1	y Hi	s As	sp A 70	la	Lys		816
-		27	15		rg CA eu Gl		28	0	u Pn	e Arg	J Le	u Ty 28	r Ty 5	T T	yr	Asn	-	864
-,	29	0 -		,p	T GC. e Al	295	5	r nei	ı ASI	n Lys	300	a Ly:	s Se	r II	Le	Val		912
30	5			u 50	A TT r Let 310	0	ı ıyı	. Met	груг	315	va.	l Phe	e Ly	s Gl	u	Lys 320		960
- 7 .			u 50	32	_	, 111L	Ser	GLY	330	Phe	Ser	· Val	l Ası	9 Ly 33	rs 1 5	Leu	1	800
_,		- 110	34	0	A TAC u Tyr	. Lys	Mec	345	Inr	Glu	Ile	Tyr	Th:	Gl	u A	Asp	1	056
7101		35	5	5 FII	r TT1 Phe	пÀг	360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Le	u A	sn	1	104
	370		, ,,,,		A TTT L Phe	375	TIE	Asn	lie	Val	380	Lys	Val	Ası	n T	yr	1:	152
385		-1-		, Gly	Phe 390	VOII	neu.	Arg	Asn	395	Asn	Leu	Ala	Ala	4	sn 00	12	00
		<b>U</b> -7	0.1.	405		GIU	116	ASI	410	Met	Asn	Phe	Thr	Lys 415	L	eu	12	48
-7-			420	Gly	TTG Leu	FIIC	GIU	425	lyr	Lys	Leu	Leu	Cys 430	Val	Aı	g	12	96
		435	****	361	AAA Lys	THE	140	ser	ren	Asp :	Lys	Gly 445	Tyr	Asn	Ly	78	13	44
	450	<i></i>	App	Dea	TGT Cys	455	rys	val	Asn	Asn :	160	Asp	Leu	Phe	Ph	e	13	92
465		-	GIU	veħ	AAT Asn 470	rne	inr .	ASD .	Asp	Leu 2 475	lsn :	Lys	Gly	Glu	Gl 48	u 0	144	10
		-	nap	485	AAT Asn	116 (	JIU 1	AIA	490	Glu G	ilu i	Asn	Ile	Ser 495	Le	น	148	8
GAT Asp	TTA Leu	ATA Ile	CAA Gln 500	CAA Gln	TAT T	TAT :	nen .	ACC : Thr 1 505	Phe	AAT I Asn P	TT (	Asp A	AAT Asn 510	GAA Glu	CC' Pr	T 0	153	6

GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 520 GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG 1632 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA 1680 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 550 555 CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA 1728 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG 1776 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA 1824 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 600 595 CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG 1872 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT 1920 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 630 635 1968 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG ATT GCA 2016 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG 2064 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA 2112 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG 2160 Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA 2208 Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG TAT AAT 2256 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 2304 Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA 2352 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile

780

785				A-677	790	. Cys	261	vai	ser	795	Leu	Met	Asn	Ser	800	2400
ATC Ile	CCT Pro	TAT	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
ACA Thr 865	TTT Phe	ACT Thr	GAA Glu	TAT Tyr	ATT Ile 870	AAG Lys	TAA •									2616

### (2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 872 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 520

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750

Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815

Asp Ala Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 860

Thr Phe Thr Glu Tyr Ile Lys \* 865 870

(2) INFORMATION FOR SEQ ID NO: 3:

WO 98/07864 PCT/GB97/02273 - 36 -

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2685 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS (B) LOCATION: 1..2685

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

	122	, 55	COTI-			\	J14	SEQ .	10 14	J: 3	:					
GGA Gly 1	TCC Ser	CCA Pro	GGA Gly	ATT Ile 5	CAT His	ATG Met	ACG Thr	TCG Ser	ACG Thr 10	CGT Arg	CTG Leu	CAG Gln	AAG Lys	CTT Leu 15	CTA Leu	48
GAA Glu	TTC Phe	GAG Glu	CTC Leu 20	CCG Pro	GGT Gly	ACC Thr	ATG Met	GAG Glu 25	TTC Phe	GTG Val	AAC Asn	AAG Lys	CAG Gln 30	TTC Phe	AAC Asn	96
TAT Tyr	AAG Lys	GAC Asp 35	CCT Pro	GTA Val	AAC Asn	GGT Gly	GTT Val 40	GAC Asp	ATT Ile	GCC Ala	TAC Tyr	ATC Ile 45	AAA Lys	ATT Ile	CCA Pro	144
AAG Lys	TAC Tyr 50	GGC Gly	CAG Gln	ATG Met	CAG Gln	CCG Pro 55	GTG Val	AAG Lys	GCT Ala	TTC Phe	AAG Lys 60	ATT Ile	CAT His	AAC Asn	AAA Lys	192
ATC Ile 65	TGG Trp	GTT Val	ATT Ile	CCG Pro	GAA Glu 70	CGC Arg	GAT Asp	ACA Thr	TTT Phe	ACG Thr 75	AAC Asn	CCG Pro	GAA Glu	GAA Glu	GGA Gly 80	240
GAC Asp	TTG Leu	AAC Asn	CCG Pro	CCG Pro 85	CCG Pro	GAA Glu	GCA Ala	AAG Lys	CAG Gln 90	GTG Val	CCA Pro	GTT Val	TCA Ser	TAC Tyr 95	TAC Tyr	288
GAT Asp	TCA Ser	ACC Thr	TAT Tyr 100	CTG Leu	AGC Ser	ACA Thr	GAC Asp	AAC Asn 105	GAG Glu	AAG Lys	GAT Asp	AAC Asn	TAC Tyr 110	CTG Leu	AAG Lys	336
GGA Gly	GTG Val	ACC Thr 115	AAA Lys	TTA Leu	TTC Phe	GAG Glu	CGT Arg 120	ATT Ile	TAT Tyr	TCC Ser	ACT Thr	GAC Asp 125	CTG Leu	GGC Gly	CGT Arg	384
ATG Met	CTG Leu 130	CTG Leu	ACC Thr	TCA Ser	ATC Ile	GTC Val 135	CGC Arg	GGA Gly	ATC Ile	CCA Pro	TTT Phe 140	TGG Trp	GGT Gly	GGC Gly	AGT Ser	432
ACC Thr 145	ATT Ile	GAC Asp	ACG Thr	GAG Glu	TTG Leu 150	AAG Lys	GTT Val	ATT Ile	GAC Asp	ACT Thr 155	AAC Asn	TGC Cys	ATT Ile	AAC Asn	GTG Val 160	480
ATC Ile	CAA Gln	CCA Pro	GAC Asp	GGT Gly 165	AGC Ser	TAC Tyr	AGA Arg	TCT Ser	GAA Glu 170	GAA Glu	CTT Leu	AAC Asn	CTC Leu	GTA Val 175	ATC Ile	528
ATC Ile	GGG Gly	CCC Pro	TCC Ser 180	GCG Ala	GAC Asp	ATT Ile	ATC Ile	CAG Gln 185	TTT Phe	GAG Glu	TGC Cys	AAG Lys	AGC Ser 190	TTT Phe	GGC Gly	576
CAC His	GAA Glu	GTG Val 195	TTG Leu	AAC Asn	CTG Leu	ACG Thr	CGT Arg 200	AAC Asn	GGT Gly	TAC Tyr	GGC Gly	TCT Ser 205	ACT Thr	CAG Gln	TAC Tyr	624

AT'	CG' Are	à bu	C AG e Se:	C CCA	A GAC	TTC Phe 215	Ini	Phe	C GGT	r Tro y Phe	GA( Glu 22(	1 GJ	G AG u Se	C CT r Le	G GAG u Glu	672
va. 229	L AS	o Ini	r Asi	ı Pro	230	Leu	GIY	' Ala	a Gly	/ Lys 235	Phe	≥ Al	a Th	r As	T CCA P Pro 240	720
Ala	va:	l Thi	c Lev	245	HIS	GIu	Leu	Ile	250	Ala	Gly	/ Hi	s Ar	25!	_	768
GIŞ	, 116	Ala	260	ASI	Pro	Asn	Arg	Val 265	Phe	Lys	Val	Ası	27(	Ası	GCC Ala	816
ıyı	Tyr	275	i Met	ser	GIY	reu	280	Val	Ser	Phe	Glu	Gl: 285	Let	Arg	ACG Thr	864
Pne	290	GIA	' H1S	Asp	Ala	Lys 295	Phe	Ile	Asp	Ser	Leu 300	Gln	Glu	Asn	GAG Glu	912.
305	Arg	Leu	Tyr	туг	310	Asn	Lys	Phe	Lys	Asp 315	Ile	Ala	Ser	Thr	CTG Leu 320	960
Asn	rys	Ala	Lys	Ser 325	Ile	Val	Gly	Thr	Thr 330	Ala	Ser	Leu	Gln	Tyr 335		1008
rys	Asn	Val	Phe 340	Lys	GIu	Lys	Tyr	Leu 345	Leu	Ser	Glu	Asp	Thr 350	Ser	-	1056
Lys	Pne	355	Val	Asp	AAA Lys	Leu	Lys 360	Phe	Asp	Lys	Leu	Tyr 365	Lys	Met	Leu	1104
Inr	370	IIe	Tyr	Thr	GAG Glu	<b>Asp</b> <b>37</b> 5	Asn	Phe	Val	Lys	Phe 380	Phe	Lys	Val	Leu	1152
385	Arg	Lys	Thr	Tyr	TTG Leu 390	Asn	Phe	Asp	Lys	Ala 395	Val	Phe	Lys	Ile	Asn 400	1200
116	Val	Pro	Lys	Val 405	AAT Asn	Tyr	Thr	Ile	Tyr 410	Asp	Gly	Phe	Asn	Leu 415	Arg	1248
Asn	Thr	Asn	<b>Leu</b> 420	Ala	GCA . Ala .	Asn :	Phe .	Asn 425	Gly	Gln .	Asn	Thr	Glu 430	Ile	Asn	1296
ASN	Met	Asn 435	Phe	Thr	AAA Lys	Leu !	Lys / 440	Asn	Phe	Thr	Gly	Leu 445	Phe	Glu	Phe	1344
Tyr	Lys 450	Leu	Leu	Cys		Arg ( 455	Gly :	Ile	Ile	Thr	Ser   460	Lys	Thr	Lys	Ser	1392
TTA Leu 465	Asp Asp	AAA Lys	GGA Gly	Tyr .	AAT AAR A	AAG ( Lys )	SCA '	ITA . Leu .	Asn .	GAT : Asp 1 475	TTA ' Leu (	TGT Cys	ATC Ile	Lys	GTT Val 480	1440

AA: Ası	r AAT n Asr	TGC	GAC Asp	TT( Lev 485	T Elif	TTT Phe	AGI Ser	CC:	TC/ Ser 490	r Gli	A GA:	T AA:	TTT n Phe	T AC	T AAT r Asn	1488
voj	, nec	. nau	500	Gly	GIL	. GIU	116	505	Ser	Asp	Thi	~ Asr	1 Ile 510	e Glu	A GCA 1 Ala	1536
nic	. GIU	515	, veri	110	. Jer	Leu	520	Leu	, 116	Gln	Glr	525	Туг	Let	ACC Thr	1584
7110	530	FILE	nsp	ASII	GIU	<b>5</b> 35	GIU	Asn	ı Ile	Ser	11e 540	Glu	Asn	Leu	TCA Ser	1632
545	Asp	116	116	GIY	550	Leu	GIU	Leu	Met	Pro 555	Asn	Ile	Glu	Arg	TTT Phe 560	1680
PIO	ASII	GIY	Lys	565	lyr	GIU	ren	Asp	Lys 570	Tyr	Thr	Met	Phe	His 575		1728
Leu	Arg	Ala	580	GIU	Pne	GIU	HIS	585	Lys	Ser	Arg	Ile	Ala 590	Leu		1776
ASII	361	595	ASII	GIU	мта	TTA Leu	600	Asn	Pro	Ser	Arg	Val 605	Tyr	Thr	Phe	1824
PILE	610	SEL	Asp	lyr	vai	AAG Lys 615	rys	Val	Asn	Lys	Ala 620	Thr	Glu	Ala	Ala	1872
625		Leu	GIY	irp	630	GAA Glu	GIn	Leu	Val	Tyr 635	Asp	Phe	Thr	Asp	Glu 640	1920
1111	Set	GIU	vai	645	inr	ACG Thr	Asp	Lys	11e 650	Ala	Asp	Ile	Thr	Ile 655	Ile	1968
116	PIO	ıyr	660	GIÀ	PIO	GCT Ala	Leu	Asn 665	Ile	Gly	Asn	Met	Leu 670	Tyr	Lys	2016
Asp	Asp	675	Val	GIÀ	АТА	TTA Leu	11e 680	Phe	Ser	Gly	Ala	Val 685	Ile	Leu	Leu	2064
GIU	690	TTE	PIO	GIU	116	GCA Ala 695	Ile	Pro	Val	Leu	Gly 700	Thr	Phe	Ala	Leu	2112
705	ser	lyr	116	ALA	710	AAG Lys	Val	Leu	Thr	Val 715	Gln	Thr	Ile	Asp	<b>Asn</b> 720	2160
Ala	Leu	ser	rys	725	ASN	GAA Glu	Lys	Trp	<b>Asp</b> 730	Glu	Val	Tyr	Lys	Tyr 735	Ile	2208
GTA Val	ACA Thr	ASI	TGG Trp 740	TTA Leu	GCA Ala	AAG Lys	Val	AAT Asn 745	ACA Thr	CAG Gln	ATT Ile	Asp	CTA Leu 750	ATA Ile	AGA Arg	2256

- 39 -

		75	5	- 010		- 266	760	)	1 GT	ı Ala	a GI	1 Ala 769	Thi	Ly	G GCT S Ala	2304
	770	)	<b>, -</b>		,-	775	GII	ı ıyı	101	GIL	780	ı Glu	Lys	Asr	AAT Asn	2352
785			•		790	nap	Leu	Ser	ser	795	Leu	Asn	Glu	Ser	ATA Ile 800	2400
	-,-			805	ASII	116	ASII	гàг	910	Leu	Asn	CAA Gln	Cys	Ser 815	Val	2448
	-,-		820		JCI	MEL	116	825	Tyr	GIA	Val	AAA Lys	Arg 830	Leu	Glu	2496
		835	7.24	561	neu	пуъ	840	Ala	Leu	Leu	Lys	TAT Tyr 845	Ile	Tyr	Asp	2544
	850	01,	••••	Deu	116	855	GIN	vai	Asp	Arg	<b>B60</b>	AAA Lys	Asp	Lys	Val	2592
865				261	870	vəħ	116	PIO	Pne	875	Leu	TCC Ser	Lys	Tyr	GTA Val 880	2640
GAT . Asp .	AAT Asn	CAA Gln	Arg	TTA Leu 885	TTA Leu	TCT . Ser	ACA Thr	Pne	ACT Thr 890	GAA Glu	TAT Tyr	ATT Ile	Lys	TAA * 895		2685

# (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 895 amino acids
  - (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu

Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn

Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys 50 55 60

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys

Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr 200 Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro 230 Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala 265 Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr 280 Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser

Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val 465 470 475 480

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 485

Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 500 505 510

Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr
515 520 525

Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 530 535

Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 555 550

Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 575

Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 580 585 590

Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 595 600 605

Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 610 615 620

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 625 630 635 640

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 645 650 655

Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
660 665 670

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 675 680 685

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 690 695 700

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 705 710 715 720

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
725 730 735

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 740 745 750

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
755 760 765

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 770 780

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 785 790 795 800

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 805 810 815 - 42 -

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 820 825 830

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 835 840 845

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 865 870 875 880

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys \* 895 890

### (2) INFORMATION FOR SEQ ID NO: 5:

#### (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2622 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:1..2622

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGA Gly 1	TCC Ser	ATG Met	GAG Glu	TTC Phe 5	GTG Val	AAC Asn	AAG Lys	CAG Gln	TTC Phe 10	AAC Asn	TAT Tyr	AAG Lys	GAC Asp	CCT Pro 15	GTA Val	48
AAC Asn	GGT Gly	GTT Val	GAC Asp 20	ATT Ile	GCC Ala	TAC Tyr	ATC Ile	AAA Lys 25	ATT Ile	CCA Pro	AAG Lys	TAC Tyr	GGC Gly 30	CAG Gln	ATG Met	96
CAG Gln	CCG Pro	GTG Val 35	AAG Lys	GCT Ala	TTC Phe	AAG Lys	ATT Ile 40	CAT His	AAC Asn	AAA Lys	ATC Ile	TGG Trp 45	GTT Val	ATT Ile	CCG Pro	144
GAA Glu	CGC Arg 50	GAT Asp	ACA Thr	TTT Phe	ACG Thr	AAC Asn 55	CCG Pro	GAA Glu	GAA Glu	GGA Gly	GAC Asp 60	TTG Leu	AAC Asn	CCG Pro	CCG Pro	192
CCG Pro 65	GAA Glu	GCA Ala	AAG Lys	CAG Gln	GTG Val 70	CCA Pro	GTT Val	TCA Ser	TAC Tyr	TAC Tyr 75	GAT Asp	TCA Ser	ACC Thr	TAT Tyr	CTG Leu 80	240
AGC Ser	ACA Thr	GAC Asp	AAC Asn	GAG Glu 85	AAG Lys	GAT Asp	AAC Asn	TAC Tyr	CTG Leu 90	AAG Lys	GGA Gly	GTG Val	ACC Thr	AAA Lys 95	TTA Leu	288
TTC Phe	GAG Glu	CGT Arg	ATT Ile 100	TAT Tyr	TCC Ser	ACT Thr	GAC Asp	CTG Leu 105	GGC Gly	CGT Arg	ATG Met	CTG Leu	CTG Leu 110	ACC Thr	TCA Ser	336
ATC Ile	GTC Val	CGC Arg 115	GGA Gly	ATC Ile	CCA Pro	TTT Phe	TGG Trp 120	GGT Gly	GGC Gly	AGT Ser	ACC Thr	ATT Ile 125	GAC Asp	ACG Thr	GAG Glu	384
TTG Leu	AAG Lys 130	GTT Val	ATT Ile	GAC Asp	ACT Thr	AAC Asn 135	TGC Cys	ATT Ile	AÁC Asn	GTG Val	ATC Ile 140	CAA Gln	CCA Pro	GAC Asp	GGT Gly	432

14	5					150	200	ns:	. De	u va	15	.e 1.	ie G	ly P	ro :	Ser	GCG Ala 160	480
GA As	C AT	T A	TC C		Phe (	GAG Glu	TGC Cys	AAC Lys	G AG	C TT r Ph 17	e GI	C CA y Hi	AC GA	AA G Lu V	al I	TTG Leu 175	AAC Asn	528
			1	80			GIY	261	185	5	n Ty	r II	T CC	g Pi	ne S 90	er	Pro	576
		19	5		-, -		<b>91</b> u	200	261	. Let	1 G11	u Va	T GA 1 As 20	p Th 5	r A	sn	Pro	624
	210	0	.,		-, -	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	215	via	Int	Asp	Pro	22 22		l Th	r L	eu	Ala	672
225	'				2	30	3 L y	uis	Arg	Leu	235	Gl	C AT	e Al	a I	le .	Asn 240	720
			<b>.</b>	2	45	ys v		ASII	Inr	250	Ala	Туз	TAC	r Gl	u Me 25	et :	Ser	768
,			26	ō		.e G	114	GIU	265	Arg	Thr	Phe	GGT Gly	/ Gly 270	/ Hi	s į	qaA	816
	-1 -	275	5		.p 00	1	2	280	GIU	ASN	Glu	Phe	CGT Arg	Leu	ту	r ]	yr	864
TAC Tyr	AAC Asn 290	AAC Lys	FT Ph	r AA e Ly	A GA 'S As	P T	TT ( le <i>1</i> 95	SCA Nla	AGT Ser	ACA Thr	CTG Leu	AAC Asn 300	AAG Lys	GCT Ala	Ly	G I s S	CC er	912
305		,			31	0	2 L T	seu (	GIN	lyr	Met 315	Lys	AAT Asn	Val	Ph	e L	ys 20	960
	-,-	-,-		32	5		iu A	ap .	inr	330	GIA	Lys		Ser	Va:	l A 5	sp	1008
AAA Lys		-,-	340	no <sub>l</sub>	y Lly	o De	u 1	yr 1	345	met	Leu	Thr	Glu	Ile 350	Туз	T	hr	1056
GAG (	- LU <sub>E</sub>	355	FILE	va.	L LJY:	9. PU	3	60 1	À8	val	Leu .	Asn	Arg 365	Lys	Thr	T	yr	1104
	370		nop	шус	, 110	37	5	ne L	ys .	ile .	Asn	11e 380	Val	Pro	Lys	Va	<b>a</b> l	1152
AAT 1 Asn 1 385	- / -		116	171	390	)	y Fi	ue A	sn I	Jeu /	Arg 2	Asn	Thr	Asn	Leu	A1	.a )0	1200
GCA A	AAC ( Asn )	TTT Phe	AAT Asn	GGT Gly 405	GIL	AA' As:	T AC	CA G	In 1	le A	AAT /	AAT Asn	ATG Met	AAT Asn	TTT Phe 415	AC Th	T	1248

AAA Lys	CTA Leu	AAA Lys	AAT Asn 420	TTT Phe	ACT Thr	GGA Gly	TTG Leu	TTT Phe 425	Glu	TTT Phe	TAT Tyr	AAG Lys	TTG Leu 430	Leu	TGT Cys	1296
GTA Val	AGA Arg	GGG Gly 435	ATA Ile	ATA Ile	ACT Thr	TCT Ser	AAA Lys 440	ACT Thr	AAA Lys	TCA Ser	TTA Leu	GAT Asp 445	AAA Lys	GGA Gly	TAC	1344
AAT Asn	AAG Lys 450	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 455	TGT Cys	ATC Ile	AAA Lys	GTT Val	AAT Asn 460	AAT Asn	TGG Trp	GAC Asp	TTG Leu	1392
TTT Phe 465	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 470	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 475	GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 480	1440
GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 485	GAT Asp	ACT Thr	TAA neA	ATA Ile	GAA Glu 490	GCA Ala	GCA Ala	GAA Glu	GAA Glu	AAT Asn 495	ATT Ile	1488
AGT Ser	TTA Leu	GAT Asp	TTA Leu 500	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 505	TTA Leu	ACC Thr	TTT Phe	AAT Asn	TTT Phe 510	GAT Asp	AAT Asn	1536
Glu	Pro	Glu 515	Asn	Ile	TCA Ser	Ile	Glu 520	Asn	Leu	Ser	Ser	Asp 525	Ile	Ile	Gly	1584
CAA Gln	TTA Leu 530	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 535	ATA Ile	GAA Glu	AGA Arg	TTT Phe	CCT Pro 540	AAT Asn	GGA Gly	AAA Lys	AAG Lys	1632
TAT Tyr 545	GAG Glu	TTA Leu	Asp	AAA Lys	TAT Tyr 550	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 555	CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 560	1680
TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 565	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 570	ACA Thr	AAT Asn	TCT Ser	GTT Val	AAC Asn 575	GAA Glu	1728
GCA Ala	TTA Leu	TTA Leu	AAT Asn 580	CCT Pro	AGT Ser	CGT Arg	GTT Val	TAT Tyr 585	ACA Thr	TTT Phe	TTT Phe	TCT Ser	TCA Ser 590	GAC Asp	TAT Tyr	1776
Val	Lys	Lys 595	Val	Asn	AAA Lys	Ala	Thr 600	Glu	Ala	Ala	Met	Phe 605	Leu	Gly	Trp	1824
GTA Val	GAA Glu 610	CAA Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 615	TTT Phe	ACC Thr	GAT Asp	GAA Glu	ACT Thr 620	AGC Ser	GAA Glu	GTA Val	AGT Ser	1872
ACT Thr 625	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 630	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 635	ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 640	1920
CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 645	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 650	AÄA Lys	GAT Asp	GAT Asp	TTT Phe	GTA Val 655	GGT Gly	1968
GCT Ala	TTA Leu	ATA Ile	TTT Phe 660	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 665	CTG Leu	TTA Leu	GAA Glu	TTT Phe	ATA Ile 670	CCA Pro	GAG Glu	2016
ATT Ile	GCA Ala	ATA Ile 675	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 680	TTT Phe	GCA Ala	CTT Leu	GTA Val	TCA Ser 685	TAT Tyr	ATT Ile	GCG Ala	2064

	69	0.				69	5	r 116	e As	P Ası	70	a Lei O	u Se:	r Ly	A AGA s Arg	211:
70	5	1	- <b></b>	P	71	0	Lly	r nys	s ryr	715	≥ Va. 5	l Thi	Ası	ı Tr	G TTA P Leu 720	2160
	,			72	5	. 116	. wai	Leu	730	Arg	; Lys	: Lys	Met	Lys 735	•	2208
			74	0	· Ale	. GIU	, WIG	745	rys	Ala	Ile	lle	750	Туг	CAG Gln	2256
-,-		75	5		. 010	Glu	760	Lys	Asn	Asn	Ile	Asn 765	Phe	Asn	ATT	2304
	770				. Ly c	CTT Leu 775	ASII	GIU	ser	He	<b>Asn</b> 780	Lys	Ala	Met	Ile	2352
785			,.		790		GIN	Cys	ser	795	Ser	Tyr	Leu	Met	Asn 800	2400
				805	GIY	GTT Val	гÀг	Arg	810	GIu	Asp	Phe	Asp	Ala 815	Ser	2448
	-,-		820	Deu	Deu	AAG Lys	TYE	825	ıyr	Asp	Asn	Arg	Gly 830	Thr	Leu	2496
	,	835	•	ASP	Arg		840	Asp	ràs	Vai	Asn	Asn 845	Thr	Leu	Ser	2544
	850			• • • • • • • • • • • • • • • • • • • •	<b>G111</b>	CTT Leu 855	361	цуs	ıyr	val.	GAT Asp 860	AAT Asn	CAA Gln	AGA Arg	TTA Leu	2592
TTA Leu 865	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 870	TAT .	ATT . Ile :	AAG : Lys	TAA *							2622

# (2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 874 amino acids
    (B) TYPE: amino acid
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val

As Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 20 25 30

Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro 35 40 45

Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly 135 Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn 170 Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp. Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro 200 Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala 215 His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 250 Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser 295 Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys-Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val 375 Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 390 395

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr 405 410 415

Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tvr Lys Leu Leu Cys
420 425 430

Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr
435
440
445

Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu 450 460

Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly 475 475 480

Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile 485 490 495

Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn 500 505 510

Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly 515 520 525

Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys 530 540

Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu 550 555 560

Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu 575

Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr 580 585 590

Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp 595 600 605

Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser 610 620

Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly 635 635

Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly 645 650

Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu 660 665 670

Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala 675 680 685

Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg 690 695 700

Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu 705 710 715 720

Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu 725 730 735

Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln 740 745 750

-	48	_

										- 48	-				
Tyr	Asn	Gln 755	Tyr	Thr	Glu	Glu	Glu 760	Lys	Asn	Asn	Ile	Asn 765	Phe	Asn	Ile
Asp	Asp 770	Leu	Ser	Ser	Lys	Leu 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile
Asn 785	Ile	Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795	Ser	Tyr	Leu	Met	Asn 800
Ser	Met	Ile	Pro	Tyr 805	Gly	Val	Lys	Arg	Leu 810	Glu	Asp	Phe	Asp	Ala 815	Ser
Leu	Lys	Asp	Ala 820	Leu	Leu	Lys	Tyr	11e 825	Tyr	Asp	Asn	Arg	Gly 830	Thr	Leu
Ile	Gly	Gln 835	Val	Asp	Arg	Leu	Lys 840	Asp	Lys	Val	Asn	Asn 845	Thr	Leu	Ser
Thr	<b>Asp</b> <b>85</b> 0	Ile	Pro	Phe	Gln	<b>Le</b> u 855	Ser	Lys	Tyr	Val	Asp 860	Asn	Gln	Arg	Leu
Leu 865	Ser	Thr	Phe	Thr	Glu 870	Tyr	Ile	Lys	* .						
(2)	INFO	RMAT	NOI	FOR	SEQ	ID N	<b>10</b> : 7	<b>'</b> :							

- (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 2613 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double

  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:

  - (A) NAME/KEY: CDS
    (B) LOCATION:1..2613

### (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATG Met 1	CCA Pro	TTT Phe	GTT Val	AAT Asn 5	AAA Lys	CAA Gln	TTT Phe	AAT Asn	TAT Tyr 10	AAA Lys	GAT Asp	CCT Pro	GTA Val	AAT Asn 15	GGT Gly	48
GTT Val	GAT Asp	ATT Ile	GCT Ala 20	TAT Tyr	ATA Ile	AAA Lys	ATT Ile	CCA Pro 25	AAT Asn	GCA Ala	GGA Gly	CAA Gln	ATG Met 30	CAA Gln	CCA Pro	96
GTA Val	AAA Lys	GCT Ala 35	TTT Phe	AAA Lys	ATT Ile	CAT His	AAT Asn 40	AAA Lys	ATA Ile	TGG Trp	GTT Val	ATT Ile 45	CCA Pro	GAA Glu	AGA Arg	. 144
GAT Asp	ACA Thr 50	TTT Phe	ACA Thr	AAT Asn	CCT Pro	GAA Glu 55	GAA Glu	GGA Gly	GAT Asp	TTA Leu	AAT Asn 60	CCA Pro	CCA Pro	CCA Pro	GAA Glu	192
GCA Ala 65	AAA Lys	CAA Gln	GTT Val	CCA Pro	GTT Val 70	TCA Ser	TAT Tyr	TAT Tyr	GAT Asp	TCA Ser 75	ACA Thr	TAT Tyr	TTA Leu	AGT Ser	ACA Thr 80	240
GAT Asp	AAT Asn	GAA Glu	AAA Lys	GAT Asp 85	AAT Asn	TAT Tyr	TTA Leu	AAG Lys	GGA Gly 90	GTT Val	ACA Thr	AAA Lys	TTA Leu	TTT Phe 95	GAG Glu	288

***	, <b></b>		10	0	I AS	h ne	u Gi	10	15 ME	sc re	eu Le	eu Ti	1: S	er 1 10	lle		336
AG( Arg	G GG G Gl	A AT y Il 11	C	A TT o Ph	T TG e Tr	G GG	r GG y Gl 12	y se	T AC	CA AT	A GA e As	AT AC Sp Th	ır G	AA 1 Lu I	TA eu	AAA Lys	384
GTT Val	11 13	c no	T AC p Th	T AA r As:	T TG n Cy	T ATT	- AS	T GT n Va	G AT	A CA e Gl	A CC n Pr 14	O As	T GC	T A y S	GT er	TAT Tyr	432
AGA Arg 145	56.	A GA	A GA	A CT	T AA' u Ası 150	T CTA n Leu 0	A GT	A AT.	A AT e Il	A GG e Gl	y Pr	C TC	A GC r Al	T G	AT sp	ATT Ile 160	480
ATA Ile	Gl:	TT:	r GAJ e Glu	A TG: 1 Cys 165	s mys	A AGO	Phe	GG/	A CA y Hi: 17	s Gl	A GT u Va	T TT l Le	G AA u As	n L	TT eu 75	ACG Thr	528
CGA Arg	AA1 Asr	GG1 Gly	TAT TY1	. Gij	C TCT	TACT Thr	CAA Glr	TAC 1 Ty:	r 110	T AGA	A TT	T AG e Se:	C CC r Pr 19	o A	AT sp	TTT Phe	576
ACA Thr	TTT	GG7 Gly 195	FILE	GAG Glu	GAG Glu	TCA Ser	CTI Leu 200	GIU	A GT	r GAT L Asp	T AC	A AAS C Asi 205	ı Pr	T CT	rT eu	TTA Leu	624
GGT Gly	GCA Ala 210	GLY	Lys	TTT Phe	GCT Ala	ACA Thr 215	GAT Asp	CCA Pro	GCA Ala	A GTA Val	ACA Thi	Lei	A GCI 1 Ala	A CA A Hi	T (	GAA Glu	672
CTT Leu 225	ATA Ile	CAT His	GCT Ala	GGA Gly	CAT His 230	AGA Arg	TTA Leu	TAT	GGA Gly	ATA Ile 235	Ala	ATT	AAT Asr	CC 1 Pr	0 1	AAT Asn 240	720
AGG Arg	GTT Val	TTT	AAA Lys	GTA Val 245	AAT Asn	ACT Thr	AAT Asn	GCC Ala	TAT Tyr 250	Tyr	GAA Glu	ATG Met	AGT Ser	GG G1 25	y I	TTA Jeu	768
GAA Glu	GTA Val	AGC Ser	TTT Phe 260	GAG Glu	GAA Glu	CTT Leu	AGA Arg	ACA Thr 265	TTT Phe	GGG Gly	GGA Gly	CAT His	GAT Asp 270	GC.	A A a I	AG ys	816
TTT Phe	ATA Ile	GAT Asp 275	AGT Ser	TTA Leu	CAG Gln	GAA Glu	AAC Asn 280	GAA Glu	TTT Phe	CGT Arg	CTA Leu	TAT Tyr 285	TAT Tyr	TA:	r A	AT sn	864
Lys .	TTT Phe 290	AAA Lys	GAT Asp	ATA Ile	GCA Ala	AGT Ser 295	ACA Thr	CTT Leu	AAT Asn	AAA Lys	GCT Ala 300	<b>AAA</b> Lys	TCA Ser	ATA	A G	TA al	912
GGT : Gly : 305	ACT Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAC Glu	L	<b>AA</b> ys 20	960
TAT (	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	L	TA eu	1008
AAA 1	Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA : Lys	Met .	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	G/A	AT Sp	1056
AAT I	-11	GTT Val 355	AAG Lys	TTT Phe	TTT . Phe	TAS A	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	<b>AAA</b> Lys	ACA Thr 365	TAT Tyr	TTG Leu	Al As	AT sn	1104

TTT Phe	GAT Asp 370	гÀг	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	Lys	GTA Val	AAT Asn	TAC	1152
ACA Thr 385	TTE	TAT Tyr	GAT Asp	GGA Gly	Phe	Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
. Phe	Asn	Gly	Gln	405		Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
Gly	Ile	11e 435	Thr	Ser	AAA Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392
AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	GAT Asp	AAT Asn 470	TTT Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	1440
Ile	Thr	Ser	Asp	Thr 485	AAT Asn	Ile	Glu	Ala	Ala 490	Glu	Glu	Asn	Ile	Ser 495	Leu	1488
Asp	Leu	Ile	Gln 500	Gln	TAT Tyr	Tyr	Leu	Thr 505	Phe	Asn	Phe	Asp	Asn 510	Glu	Pro	1536
GAA Glu	AAT Asn	ATT Ile 515	TCA Ser	ATA Ile	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	AGT Ser	GAC Asp	ATT Ile	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu	1584
Glu	<b>Leu</b> 530	Met	Pro	Asn	ATA Ile	Glu 535	Arg	Phe	Pro	Asn	Gly 540	Lys	Lys	Tyr	Glu	1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1824
CAA Gln	TTA Leu 10	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	Ąsp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	1920

				64	5	G TT! t Lei	1 1 Y	L Ly.	650	P As <sub>1</sub>	p Pho	e Val	. Gly	y Al 65	a L	eu	1968
			66	0		T ATT	: Let	669	1 GI	ı Phe	e Ile	Pro	670	ı Il	еA	la	2016
		67	5	u	y 141.	T TTT r Phe	680	Let	ı val	. Ser	r Tyr	Ile 685	Ala	As:	n L	ys	2064
GT Va	T CT		C GT r Va	T CA 1 Gl	A ACI n Thi	A ATA Tle 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	A AGT Ser 700	Lys	AGA Arg	AA' Ası	T G)	AA Lu	2112
<b>AA</b> Ly: 70:		G GA	T GA	G GT u Va	TAT 1 Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	Asn	TGG Trp	TTA Leu	GCA Ala	A AA A Ly 72	'S	2160
GT: Va	TAAT LAST	AC.	A CAG	G ATT	- veh	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	۱ Le	'A ·u	2208
GA/ Glu	AAT AST	CA Gli	A GCA n Ala 740	910	A GCA 1 Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	' AA ' As	T n	2256
CAA Glm	TAT	ACT Thi 755	GIU	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GA' As	T P	2304
TTA Leu	AGT Ser 770	TCC	E AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	AT	A e	2352
AAT Asn 785	AAA Lys	TTT	TTG	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	Ser	TAT Tyr 795	TTA Leu	ATG : Met :	AAT Asn	TCT Ser	ATO Met 800	:	2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA ( Leu (	GTI	GAT Asp 810	TTT Phe	GAT Asp	GCT / Ala S	Ser :	CTT Leu 815	AAA Lys	<b>.</b>	2448
GAT Asp	GCA Ala	TTA Leu	Tra Leu 820	AAG Lys	TAT Tyr	ATA :	rar 1	GAT Asp 825	AAT . Asn .	AGA Arg	GGA . Gly :	Thr I	TTA I	ATT Ile	GGT Gly	•	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT A	AAA ( Lys 1 340	GTT . Val :	AAT I Asn I	AAT : Asn :	Thr 1	CTT A Leu S 345	GT # Ser 1	ACA Thr	GAT Asp		2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	Ser	AAA 1 Lys 1 855	TAC ( Tyr \	GTA (	GAT / Asp /	Asn (	CAA 1 Gln 1 860	AGA T Arg L	TA I	TA Seu	TCT Ser		2592
ACA Thr 865	TTT Phe	ACT Thr	GAA Glu	TAT Tyr	ATT . Ile : 870	AAG Lys											2613

# (2) INFORMATION FOR SEQ ID NO: 8:

<sup>(</sup>i) SEQUENCE CHARACTERISTICS:

<sup>(</sup>A) LENGTH: 871 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

- 52 -

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Île Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435
440
445

Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 450 460

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480

Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495

Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 510

Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
530 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala
660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685 - 54 -

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750

Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp
755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860

Thr Phe Thr Glu Tyr Ile Lys 865 870

#### (2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2628 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:1..2628
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT 48
Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20
25

GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA 192
Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
50 60

65	5		va.	I FL	70	)	ı ıy	r ly	I AS	p se: 7:	r Th	т Ту	r Le	u Se	C ACA Thr	•
Val	, ve		u by.	85	5	. Iyi	re	и шу	9 GI	y Val	. Thi	r Ly:	s Le	u Ph	C GAG e Glu 5	
Arg	1116	z Ly	100	)	. Asp	, rec	i Gi)	10:	g met 5	Leu	Let	ı Thi	Se:	r Il	C GTC e Val	
CGC Arg	GG# Gly	A ATO 116	PIC	A TTI	TGG Trp	GGT Gly	GG( Gly 120	Sei	r Acc	: ATT	GAC Asp	ACG Thr 125	Glı	TT Le	G AAG u Lys	384
GTT Val	Ile 130	. vol	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC	GTC Val	ATC L Ile	CAA Gln	CCA Pro 140	Asp	GGT Gly	AG Se:	TAC Tyr	432
145	261	GIU	GIU	Leu	150	ren	Vai	Ile	: Ile	Gly 155	Pro	Ser	Ala	Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTC Let 175	ACG Thr	528
CGT	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GIY	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	ALA	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC (	GTG Val	TTC Phe	гÀв	GTT Val 245	ABI	ACC Thr	AAC Asn	Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA ( Glu '	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA ( Glu )	CTG ( Leu .	arg	ACG Thr 265	TTC Phe	GGT (	GGC Gly	His .	GAT Asp 270	GCG Ala	AAG Lys	816
TTT I	176	GAC Asp 275	AGC Ser	TTG Leu	CAG ( Gln (	stu 1	AAC Asn 280	<b>GAG</b> Glu	TTC Phe	CGT ( Arg )	Leu '	TAC ' Tyr ' 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG 1	Phe	AAA Lys	GAT . Asp	ATT (	ALA S	AGT 1 Ser 1 195	ACA Thr	CTG . Leu .	AAC / Asn :	Lys )	SCT A Ala 1 300	AAG : Lys :	rcc . Ser	ATT Ile	GTG Val	912
GGT A Gly 1 305	ACC I	ACT (	GCT ( Ala (	ser i	TTA C Leu G 310	CAG 1	rat : Cyr !	ATG . Met :	Lys I	AAT ( Asn V	at :	Phe I	AAA ( Lys (	GAG Glu	AAA Lys 320	960
TAT C	TC ( Leu )	CTA ' Leu :	ser (	GAA ( Glu / 325	ATA Asp I	CA 7	CT ( Ser (	Gly :	AAA 1 Lys 1 330	Phe S	CG ( Ser V	GTA ( /al /	/sp 1	AAA Lys 335	TTA Leu	1008

•			34	0	,	. <b>.</b> .	o Met	34	u 111.	r GI	u Il	е Ту	r Th	r Gl 0	G GAT u Asp	
,,,,,		35	5		- 111	- Uys	360	)	J ASI	n Arg	g Ly:	s Th 36	r Ту 5	r Le	G AAT u Asn	1104
	370	)	, ,,_,			375	116	ASI	1 116	• Val	380	b Ly:	s Va	l As	T TAC	1152
385		,.		, 01,	390	Yall	Leu	Arg	AST	395	Asr	ı Lei	ı Ala	a Ala	A AAC A Asn 400	1200
		. 017	<b>U</b> 11.	405		GIU	116	ASI	410	Met	Asn	Phe	Thi	419		1248
-, -			420	Gry	Deu	FIIE	GIU	425	Tyr	Lys	Leu	Leu	430	Va]	A AGA L Arg	1296
,		435	****	Jer	БУБ	1111	440	ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	AAG Lys	1344
-	450	nap	GIY	ALG	Den	455	АБР	Leu	Cys	Ile	Lys 460	Val	Asn	Asn	TGG Trp	1392
465	500		rne	361	470	261	GIU	Asp	Asn	Phe 475	Thr	Asn	Asp	Leu	AAT Asn 480	1440
-, -	O <sub>2</sub> y	0.1	Jiu	485	****	361	wab	inr	490	He	Glu	Ala	Ala	Glu 495		1488
	***	061	500	vaħ	TTA Leu	116	GTÜ	505	Tyr	Tyr	Leu	Thr	Phe 510	Asn	Phe	1536
p		515	110	Giu	AAT Asn	116	520	116	GIu	Asn	Leu	Ser 525	Ser	Asp	Ile	1584
110	530	GIM	neu	GIU		535	PIO .	asn	Ile	Glu	Arg 540	Phe	Pro	Asn	Gly	1632
545	<b>.</b> ,,,	-7-	GIU	neu	GAT Asp 550	гуs	ıyr	Inr	Met	Phe 555	His	Tyr	Leu	Arg	<b>Ala</b> 560	1680
<b>J111</b>	JIL	FIIC	GIU	565	GGT . Gly	Lys .	ser,	Arg	570	Ala	Leu	Thr	Asn	Ser 575	Val	1728
7011	914	A14	580	neu .	AAT Asn	PIO :	ser /	885	Val	Tyr '	Thr	Phe	Phe 590	Ser	Ser	1776
GAC Asp	TAT	GTA Val 595	AAG Lys	AAA Lys	GTT /	ASN .	AAA ( Lys ) 600	GCT . Ala '	ACG (	GAG ( Glu /	Ala .	GCT Ala 605	ATG Met	TTT Phe	TTA Leu	1824

GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu 610 620	1872
GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 625 630 640	1920
ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT	1968
GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile 660 665 670	2016
CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr 685	2064
ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser 690 695 700	2112
AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn 715 720	2160
TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met 725 730 735	2208
AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn 740 745	2256
TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn Ile Asn Phe 760 765	2304
AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala 770 775 780	2352
ATG ATT AAT ATA AAT AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu 790 795 800	2400
ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp 805 810 815	2448
GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly 820 825	2496
ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT ACA Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr 835 840 845	2544
CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln 850 860	2592
AGA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys + 865 870 875	2628

### (2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 876 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435
440
445

Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp
450 455 460

Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn 465 470 475 480

Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 485

Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 500 510

Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile
515 520 525

Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly 530 540

Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala 545 550 560

Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val

Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser 585

Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu 595 605

Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu
610 620

Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 625 630 635 640

Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 655

WO 98/07864 PCT/GB97/02273

- 60 -

 Val
 Gly
 Ala
 Leu 660
 Ile 660
 Free 660
 Ser Gly
 Ala 665
 Val
 Ile Leu 665
 Leu 670
 Phe Ile 670

 Pro
 Glu
 Ile Ala Asa Ile Asa Ile Pro
 Val Leu 680
 Gly Thr Phe Ala Leu Leu 685
 Val Ser Tyr 685

 Ile Ala Asa Asa Lys
 Val Leu Thr 695
 Val Gln Thr Ile Asa Asa Asa Ala Leu Ser 700
 Asa Ala Leu Ser 700

 Lys Arg Asa Asa Glu Lys Trp Asa Glu Lys Trp Asa Glu Val Tyr Lys Tyr Ile Val Thr 720
 Trp Leu Ala Lys Val Asa Thr Glu Ile Asa Leu 1 Lys Asa Asa Ile Asa 725
 Asa Thr Glu Asa Thr Lys Asa Ile Asa Ile Asa 725
 Ile Asa 735
 Ile Asa 750
 Ile Asa Phe 735

 Lys Glu Ala Leu Glu Asa Leu Glu Asa Gla Tyr Thr Glu Glu Glu Lys Asa Asa Asa Ile Asa Phe 760
 Free Lys Leu Asa Glu Lys Asa Asa Ile Asa Lys Ala 765
 Ile Asa Phe 765
 Ile Asa Phe 765

 Asa Ile Asa Ile Asa Leu Ser Lys Leu Asa Glu Ser Ile Asa Lys Ala 770
 Free Ile Asa Phe 876
 Ile Asa Phe 880
 Ile Asa Phe 881

 Met Ile Asa Ser Met Ile Pro Tyr Gly Val Lys Tyr Ile Tyr Asa Asa Arg Gly 882
 Free Ile Tyr Asa Asa Arg Gly 882
 Ala Leu Lys Asa Asa Tyr Gly 883

 Ala Ser Leu Lys Asa Asa Arg Leu Lys Asa Lys Tyr Val Asa Asa Asa Thr 883
 Free Asa Sa Gly 884

Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 865 870 875

### (2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2637 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE\_TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:1..2637
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
- ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 15

  GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 25

	- <b></b> _	35		AAG Lys		*****	40	пуs	> 11	e Ti	.pv	al I	le P 45	ro (	3lu	Arg	144
	50			AAC Asn		55	GIU	GIY	AS	р ге	u As	in Pi	ro P	ro F	ro	Glu	192
65	2,0	<b></b>	-	CCA Pro	70	Ser	TYL	ıyı	ASI	9 Se 7	r 11. 5	r Ty	T L	eu S	er	Thr 80	240
	*.0	014	<b>5</b> ,5	GAT Asp 85		TYL	Den	Lys	90	va )	1 Th	r Ly	's L€	eu P	he 95	Glu	288
5		-,-	100	ACT Thr	nop	DCG .	GIY	105	met	Lei	ı Le	u Th	r Se	r II	le	Val	336
5	,	115		TTT 'Phe'	p	Gly (	120	ser	Inr	. 116	As)	P Th	r Gl 5	u Le	u :	Lys	384
	130		••••	AAC : Asn (	-ys :	135	4811	vaı	TIE	GII	140	) )	Gl	y Se	r :	ſyr	432
145			olu .		50	seu v	al.	rie	TTE	155	Pro	Sez	. Ala	a As	p ] 1	le .60	480
				IGC A Cys I 165	ya s	er r	ne (	ыу	170	Glu	Val	Leu	Asr	1 Le	u 1 5	hr	528
3		,	180	GC T			111 1	. <b>8</b> 5	TTE	Arg	Phe	Ser	190	Ası	P	he	576
ACG :	1	95		114 0	.u	2	00	ıu v	vaı	Asp	Thr	Asn 205	Pro	Leu	ı L	eu	624
=	10	, -	.,		2	15	ap P	IO A	на	Val	220	Leu	Ala	His	G.	lu	672
CTG A Leu I 225			<b></b>	23	0	.g De	su 1	yr G	ily	235	Ala	Ile	Asn	Pro	A:	3n 10	720
CGC G Arg V			2	45	11	IL AS	n A.	1a 1 2	50	lyr	Glu	Met	Ser	Gly 255	Le	:u	768
GAA G Glu V		2	60	Lu Gi	u De	u Ar	26	ir P 55	ne c	31Å	Gly	His	Asp 270	Ala	Ly	<b>'</b> 8	816
TTT A	2	75		su Gi	GI	28	0	u P	ne A	arg :	Leu	Tyr 285	Tyr	Tyr	As	n	864
AAG T Lys Pl 29	TT AL ne Ly 90	AA GA	AT AT	T GC e Al	A AG a Se 29	LIL	A CI	G A	AC A sn L	ys A	SCT Ala	AAG Lys	TCC Ser	ATT Ile	GT Va	G 1	912

GG: Gl; 305		C AC	C GCT	TCI Sei	Leu 310	GIL	TAT	C ATC	G AAA	AAT Asn 315	ı Val	r TT? L Phe	r aa. ⊇ Ly:	A GA	G AAA u Lys 320	960
TA1	CTC	CTA Lev	A TCT 1 Sex	GAA Glu 325	. Asp	ACA Thr	TCT Ser	GG/ Gly	A AAA / Lys 330	Phe	TCC Ser	GTA Val	A GA	T AAI D Lys 335	A TTA 5 Leu 5	1008
AAA Lys	TTI Phe	GAT Asp	Lys 340	ned	TAC	Lys	ATC Met	Leu 345	Thr	GAG Glu	ATT	TAC Tyr	ACI Thi	: Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	. Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	Ty	TTC Lev	AAT Asn	1104
TTT Phe	GAT Asp 370	nys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	Lys	GTA Val	AAT Asn	TAC	1152
ACA Thr 385		TAT	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT	ACT Thr 420	GGA Gly	TTG Leu	TTT	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	TAA Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	TAA Asn	GAA Glu	Pro	GAA Glu 520	AAT Asn	ATT Ile	TCA . Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GCC	GIN .	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	Pro .	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632
CCT Pr 545	AAT Asn	GGA Gly	AAA Lys	AAG Lys	TAT ( Tyr ( 550	GAG '	TTA Leu	GAT Asp	Lys	TAT . Tyr '	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 560	1680
CTT Leu	CGT Arg	GCT Ala	GIII	GAA Glu 565	TTT (	GAA Glu	CAT His	Gly	AAA Lys 570	TCT /	AGG Arg	ATT	GCT Ala	TTA Leu 575	ACA Thr	1728

AAT TOT GTT ANG CAA CGA TOO		
AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr 580 585 590	Thr Phe	1776
TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu 595 600 605	Ala Ala	1824
ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr 610 620	Asp Glu	1872
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT I Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr 1 635	Ile Ile 640	1920
	Tyr Lys 555	1968
GAT GAT TTT GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT C Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile L 660 665	eu Leu	2016
GAA TTT ATA CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT G Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe A 675 680 685	la Leu	2064
GTA TCA TAT ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA G Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile A 690 700	sp Asn	2112
GCT TTA AGT AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TA Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Ty 705 710 715	/r Ile 720	2160
GTA ACA AAT TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA AT Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Il 725 730 73	.e Arg 5	2208
AAA AAA ATG AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AA Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Ly 740 745 750	s Ala	2256
ATA ATA AAC TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AA Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys As: 755 760 765	n Asn	2304
ATT AAT TTT AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT TILE Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser 770 775	r Ile	2352
AAT AAA GCT ATG ATT AAT AAA AAT AAA TTT TTG AAT CAA TGC TCT Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser 785 790 795	Val 800	2400
TCA TAT TTA ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu 805 810	Glu	2448
GAT TTT GAT GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr 820 825	. Yab	2496
AAT AGA GGA ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys 835 840 845	GTT Val	2544

- 64 -

AAT AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA 2592
Asn Asn Asn CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA ASP Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys +

- (2) INFORMATION FOR SEQ ID NO: 12:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 879 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

- Glu Val. Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270
- Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285
- Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300
- Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320
- Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335
- Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350
- Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355
- Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380
- Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395
- Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415
- Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430
- Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
  435
  440
  445
- Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460
- Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480
- Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495
- Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510
- Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515
- Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540
- Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 555 560
- Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575
- Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590
- Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605

- Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620
- Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640
- Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
- Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670
- Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685
- Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700
- Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
  705 710 715 720
- Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
  725 730 735
- Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
  740 745 750
- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
  770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 865 870 875
- (2) INFORMATION FOR SEQ ID NO: 13:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2862 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (ix) FEATURE:
    - (A) NAME/KEY: CDS
    - (B) LOCATION:1..2862
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

1 5	-yo oin rhe Asi	TAT AAG GAC CCT GTA AAC GGT Tyr Lys Asp Pro Val Asn Gly 10	48
GTT GAC ATT GCC TAC Val Asp Ile Ala Tyr	ATC AAA ATT CCA Ile Lys Ile Pro 25	AAC GCC GGC CAG ATG CAG CCG Asn Ala Gly Gln Met Gln Pro 30	96
35	40	ATC TGG GTT ATT CCG GAA CGC Ile Trp Val Ile Pro Glu Arg 45	144
50	55	GAC TTG AAC CCG CCG CCG GAA Asp Leu Asn Pro Pro Pro Glu 60	192
65	70 Tyl Tyl	GAT TCA ACC TAT CTG AGC ACA Asp Ser Thr Tyr Leu Ser Thr 75 80	240
85	on lyr ned hys	GGA GTG ACC AAA TTA TTC GAG Gly Val Thr Lys Leu Phe Glu 90 95	288
100	105	ATG CTG CTG ACC TCA ATC GTC Met Leu Leu Thr Ser Ile Val 110	336
115	120	ACC ATT GAC ACG GAG TTG AAG Thr Ile Asp Thr Glu Leu Lys 125	384
130	135	ATC CAA CCA GAC GGT AGC TAC Ile Gln Pro Asp Gly Ser Tyr 140	432
145 15	o	ATC GGG CCC TCC GCG GAC ATT lle Gly Pro Ser Ala Asp Ile 155	480
165	1 oct the Gly h	CAC GAA GTG TTG AAC CTG ACG lis Glu Val Leu Asn Leu Thr 70 175	528
180	185	TT CGT TTC AGC CCA GAC TTC le Arg Phe Ser Pro Asp Phe 190	576
195	200	TT GAT ACC AAC CCG CTG TTG al Asp Thr Asn Pro Leu Leu 205	624
210	215	CG GTG ACC CTG GCA CAC GAG la Val Thr Leu Ala His Glu 220	672
225 230	And new tyr G.	GC ATT GCG ATT AAC CCG AAC ly Ile Ala Ile Asn Pro Asn 235 240	720
245	25	433	768
GAA GTA AGC TTC GAG GAA Glu Val Ser Phe Glu Glu 260	CTG CGC ACG TT Leu Arg Thr Ph 265	C GGT GGC CAT GAT GCG AAG ee Gly Gly His Asp Ala Lys 270	816

TTT Phe	ATC	GAC Asp 275	ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	Tyr	TAC	AAC Asn	864
AAG Lys	TTT Phe 290	гÀв	GAT Asp	ATT	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC	ATT Ile	GTG Val	912
305	inr	Thr	Ala	Ser	310	Gln	Tyr	Met	Lys	Asn 315	Val	Phe	Lys	Glu	320	960
TYE	Leu	ren	ser	325	Asp	Thr	Ser	Gly	1330	Phe	Ser	Val	Asp	Lys 335		1008
AAA Lys	TTT	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
Pne	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
Phe	Asn	GIA	Gln	Asn 405	Thr	GAA Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
TTT Phe	TAA Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA Glu	CCT Pro	GAA Glu 520	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632

5	45		•	AAA A Lys I	5	50	u .	,cu ,	,ab	гÀЯ	55!	r r. 5	nr M	let )	Phe	His	Tyr 560	1680
		•			65			.15	, r y	570	Sei	r Az	rg I	le A	la:	Leu 575	Thr	1728
			5	AC G sn G 80			cu D	5	85	PIO	Ser	: Az	rg V	al T 5	yr 1 90	Chr	Phe	1776
		5	95	AC T.			6	00	aı /	4511	гÀЗ	A A I	a TI 60	nr G D5	lu A	lla	Ala	1824
	61	.0		GC TO ly T		6:	15	r11 D	eu (	/aı	ıyr	AS 62	p Pr O	ne T	hr A	sp	Glu	1872
62	5			TA AC al Se	63	ō	it As	ים עי	/S 1	те	635	As	p Il	e Ti	nr I	le	Ile 640	1920
		•		TA GG le Gl 64	5		a De	u As	6	50	GIĀ	Ası	n Me	t Le	u T	yr 1 55	Lys	1968
•		-	66		,		u 11	66	5	er (	GIĀ	Ala	ı Va	1 I1 67	e Le	eu I	Leu	2016
		67	5	A GA			68	0	o va	at 1	Leu	Gly	7 Th:	r Ph 5	e Al	a I	eu	2064
	690	)		T GCC e Ala		69	5	r ne	u ir	ir v	/al	G1n 700	Thi	r Il	e As	p A	sn	2112
705			, -	A AGA	710	. 010	ı Dys	, TIE	AS	1 <b>p</b> G	15 '15	Val	Tyr	Lys	ту	r I 7	le 20	2160
				725		- Ly c	, vai	wat	73	T G	ın :	He	Asp	Lev	1 Ile 73	e A: 5	rg	2208
•	•		740			200	GIU	745	GI.	n A	ıa G	stu	Ala	Thr 750	Lys	3 A.	La	2256
		755	-,-	CAG Gln	-,-		760	ıyı	I'A;	r G.	Lu G	Hu	Glu 765	Lys	Asr	1 As	in	2304
	770			ATT Ile		775	neu	Ser	561	c L)	78 L 7	80	Asn	Glu	Ser	· Il	e	2352
785	-4 -			ATT Ile	790	***	Ven	rys	PRE	79	eu A	sn	Gln	Cys	Ser	Va 80	1 0	2400
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAI Tyr 810	. GI	T G y V	TT al	AAA Lys	CGG Arg	TTA Leu 815	GA Gl	A u	2448

GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT	ATA Ile 830	TAT Tyr	GAT Asp		2496
AAT Asn	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val		2544
AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val		2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880		2640
CCT Pro	GGA Gly	CCG Pro	GAG Glu	ACG Thr 885	CTC Leu	TGC Cys	GGG Gly	GCT Ala	GAG Glu 890	CTG Leu	GTG Val	GAT Asp	GCT Ala	CTT Leu 895	CAG Gln	:	2688
TTC Phe	GTG Val	TGT Cys	GGA Gly 900	GAC Asp	AGG Arg	GGC Gly	TTT Phe	TAT Tyr 905	TTC Phe	AAC Asn	AAG Lys	CCC Pro	ACA Thr 910	GGG Gly	TAT Tyr	;	2736
GGC Gly	TCC Ser	AGC Ser 915	AGT Ser	CGG Arg	AGG Arg	GCG Ala	CCT Pro 920	CAG Gln	ACA Thr	GGT Gly	ATC Ile	GTG Val 925	GAT Asp	GAG Glu	TGC Cys	:	2784
TGC Cys	TTC Phe 930	CGG Arg	AGC Ser	TGT Cys	GAT Asp	CTA Leu 935	AGG Arg	AGG Arg	CTG Leu	GAG Glu	ATG Met 940	TAT Tyr	TGC Cys	GCA Ala	CCC Pro	ā	2832
CTC Leu 945	AAG Lys	CCT Pro	GCC Ala	AAG Lys	TCA Ser 950	GCT Ala	GAA Glu	GCT Ala	TAG *							2	2862

### (2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 954 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

- Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
- Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140
- Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
- Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175
- Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190
- Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205
- Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220
- Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 235 240
- Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255
- Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270
- Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285
- Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300
- Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315
- Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335
- Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 350
- Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365
- Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 380
- Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400
- Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415
- Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430
- Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445
- Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460

				Leu	4,0					475	•				480
				Gly 485					491	J				495	5
								303	,				510	)	
Phe	Asn	Phe 515	Asp	Asn	Glu	Pro	Glu 520	Asn	Ile	e Ser	: Ile	Glu 525	ı Asr	Lev	Ser
Ser	Asp 530	Ile	Ile	Gly	Gln	Leu 535	Glu	Leu	Met	Pro	Asn 540	ı Ile	Glu	Arg	Phe
Pro 545	Asn	Gly	Lys	Lys	Tyr 550	Glu	Leu	Asp	Lys	Tyr 555	Thr	Met	Phe	His	Tyr 560
Leu	Arg	Ala	Gln	Glu 565	Phe	Glu	His	Gly	Lys 570	Ser	Arg	Ile	Ala	Leu 575	Thr
Asn	Ser	Val	Asn 580	Glu	Ala	Leu	Leu	<b>Asn</b> 585	Pro	Ser	Arg	Val	Tyr 590	Thr	Phe
Phe	Ser	Ser 595	Asp	Tyr	Val	Lys	Lys 600	Val	Asn	Lys	Aļa	Thr 605	Glu	Ala	Ala
Met	Phe 610	Leu	Gly	Trp ·	Val	Glu 615	Gln	Leu	Val	Tyr	Asp 620	Phe	Thr	Asp	Glu
Thr 625	Ser	Glu	Val	Ser	Thr 630	Thr	Asp	Lys	Ile	Ala 635	Asp	Ile	Thr	Ile	Ile 640
Ile	Pro	Tyr	Ile	Gly 645	Pro	Ala	Leu	Asn	Ile 650	Gly	Asn	Met	Leu	Tyr 655	Lys
Asp	Asp	Phe	Val 660	Gly .	Ala	Leu	Ile <sub>.</sub>	Phe 665	Ser	Gly	Ala	Val	Ile 670	Leu	Leu
Glu	Phe	Ile 675	Pro	Glu :	Ile .	Ala	Ile 680	Pro	Val	Leu	Gly	Thr 685	Phe	Ala	Leu
Val	Ser 690	Tyr	Ile	Ala i	Asn :	Lys ' 695	Val	Leu	Thr	Val	Gln 700	Thr	Ile	Asp	Asn
Ala 705	Leu	Ser	Lys .	Arg /	Asn ( 710	Glu 1	Lys '	Trp	Asp	Glu 715	Val	Tyr	Lys	Tyr	Ile 720
Val	Thr	Asn	Trp	Leu / 725	Ala 1	Lys \	/al /	Asn	Thr 730	Gln	Ile	Asp	Leu	Ile 735	Arg
Lys	Lys	Met	Lys (	Glu A	Ala 1	Leu (	3lu į	Asn 745	Gln	Ala	Glu	Ala	Thr 750	Lÿs	Ala
Ile	Ile	Asn 755	Tyr (	Gln 1	yr I	Asn C	3ln :	Γyr '	Thr	Glu	Glu	Glu 765	Lys	Asn	Asn
Ile .	Asn 770	Phe .	Asn :	Ile A	sp A	Asp I 775	Leu s	Ser :	Ser	Lys	Leu 780	Asn	Glu	Ser	Ile
Asn : 785	Lys	Ala	Met :	Ile A	sn 1 '90	le A	lsn I	Lys	Phe	Leu . 795	Asn	Gln	Cys		Val 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810

A	sp P	he i	Asp i	Ala 9 320	Ser	Leu	Lys	Asp	Ala 825	Leu	Leu	Lys	туг	: Ile 836		r Ası	p
A	sn A	rg (	31y 7	Thr I	Leu :	Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys 845	Ası		/s Val	L
A	sn A	sn 1 50	hr I	eu S	er 1	Thr	Asp 855	Ile	Pro	Phe	Gln	Leu 860	Ser	Lys	ту	r Val	
											8/5					r Arg 880	
				_						990					89		
			_						303.					910		y Tyr	
			_					720					925			ı Cys	
						•					Glu	Met 940	Tyr	Cys	Ala	a Pro	
Le 94	u Ly 5	s Pı	O A	la Ly	/s Se 9!	er A 50	la (	Glu .	Ala	*							
(2	) IN	FORM	MTIC	N FC	R SI	EQ I	D NO	): 1	5 :								
		c) F	(B) (C) (D) OLEC EATU (A)	LENG TYPE STRA TOPO ULE RE: NAME LOCA	: nu NDEI LOGY TYPE	ncle ONES : 1 : Di	ic a S: d inea NA (	icid loubl r geno	le								
				NCE I													
ATG Met 1	CAG Gln	TT(	GTC Val	AAC L Asi	C AA	G CA s Gl	IG T	TC A	sn T	AT A yr L 10	AG G	AC C	CT G	TA A	AAC Asn 15	GGT Gly	48
GTT																	
			20	)		- <b>L</b> y	5 1.	.e P	25	sn A	CC GG	ly G	ln M	et G 30	AG Sln	Pro	96
GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT	CA Hi	T AF	AC AI	25 AA A'	SN A TC T( le T:	Ia G. GG G: rp Va	TT A	In M TT C le P:	et G 30 CG G ro G	AG In	Pro CGC Arg	96
GTG Val GAT Asp	AAG Lys ACA Thr	GCT Ala 35 TTT Phe	TTC Phe ACG	AAC Asn	ATTI Ile	CA Hi GA G1	T AFS AS	AC AI In Ly IO	AA A' ys I: GA GJ	SR A	GG GT rp Va rG AA	TT A'AL I	In M TT Colle P:	et G 30 CG G ro G	AA lu CG	Pro CGC Arg GAA Glu	
GTG Val GAT Asp GCA Ala 65	AAG Lys ACA Thr 50 AAG Lys	GCT Ala 35 TTT Phe CAG Gln	20 Phe ACG Thr	AAC ASD CCA	CCG Pro	GA GA GI S TC	T AF S AS A GA U G1 5 A TA T Ty	AC AMIN Ly O A GO A GO A GO TA TY	AA AY YB II GA GA IY As	SR A  TC TC  Le T:  AC TC  AC TC  AC TC  AC TC  AC TC	GG GT TP Va TG AA	TT AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	In M TT Colle P: 45 CG Coro P: AT CT	et G 30 CG G ro G CG C ro P CG A	CAG	Pro CGC Arg GAA Glu ACA Thr 80	144

CG Ar	T AT	T TA	T TC r Se		T GAG	C CTC	G GG(	C CG: / Arc	g Me	G CTO	G CT	G ACC	C TC	r Il	C GTC e Val	336
CGG	G GG/ G Gly	A ATO		A TT	r TG( e Trį	G GG7	GG( Gl)	Sei	r ACC	E ATT	GA(	C ACC	GAC		G AAG 1 Lys	384
GT Val	T ATT		C ACT	r Aac r Asi	TGC Cys	TATE The	MSI	GTC Val	S ATO	CAA Glr	CCA Pro	) Asp	GG1 Gly	AG( Se)	TAC Tyr	432
145	, 501		. 010	. Dec	150	Leu	val	TTE	: Ile	155	Pro	Ser	Ala	Asp	ATT Ile 160	480
	. 011		. 910	165	Lys	ser	Pne	GIA	170	Glu	Val	. Leu	Asn	Leu 175		528
3	7.014	Cly	180	Gly	261	Inf	GIN	185	lle	Arg	Phe	Ser	Pro 190	Asp	TTC Phe	576
		195	2116	Giu	GIU	Sel	200	GIU	vaı	Asp	Thr	Asn 205	Pro	Leu		624
<b>41</b>	210	Gly	цуэ	FILE	AId	215	Asp	Pro	Ala	Val	Thr 220	CTG Leu	Ala	His	Glu	672
225	116	*****	n.a	GIY	230	Arg	Leu	Tyr	GIY	11e 235	Ala	ATT Ile	Asn	Pro	Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GIU	val	ser	260	GIU	GIU	Leu	Arg	Thr 265	Phe	Gly	Gly	CAT His	Asp 270	Ala	Lys	816
FIIC	116	275	ser	Leu	GIN	GIU	280	Glu	Phe	Arg	Leu	TAC Tyr 285	Tyr	Tyr	Asn	864
пув	290	rys	vab	TTE	AIA	295	Thr	Leu	Asn	Lys	Ala 300	AAG Lys	Ser	Ile	Val	912
305	Int	ınr	WIG	Ser	310	GIN	Tyr	Met	Lys	Asn 315	Val	TTT Phe	Lys	Glu	Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GLY	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
гув	Pne	Asp	340	ren	Tyr	rys	Met	Leu 345	Thr	Glu	Ile		Thr 350	Glu	Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	rys	GTA Val 360	CTT . Leu .	AAC Asn	AGA . Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104

TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT T Phe Asp Lys Ala Val Ph Lys Ile Asn Ile Val Pro Lys Val Asn T 370 375 380	AC 1152 yr
	sn 00
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA CT Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Le 405 410 415	eu
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AG Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Ar 420 425	g
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AA Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Ly 445	s
ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GT Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Va. 450 455 460	1
AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAA Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Ass 465 470 475	n O
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495	1
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505	•
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525	
AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 540	1632
CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 560	1680
CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACA Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 575	1728
AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	1776
TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	1824
ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA  Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu  610 620	1872
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640	1920

ATT Ile	CCA Pro	TAT	ATA	GGA Gly 645	Pro	GCT Ala	TTA Leu	AA1 Asn	TATA	Gly	AAT Asn	ATC Met	TTA Leu	TAT Tyr 655	AAA Lys	1968
GAT Asp	GAT Asp	TTI Ph	GTA Val 660	GIA	GCT Ala	TTA Leu	ATA Ile	Phe	Ser	GGA Gly	GCT Ala	GTT Val	ATI Ile 670	Leu	TTA Leu	2016
GAA Glu	TTT	Ile 675	Pro	GAG Glu	ATT	GCA Ala	ATA Ile 680	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT	2064
GTA Val	TCA Ser 690	Tyr	ATT	GCG Ala	AAT Asn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	ACA Thr	ATA	GAT Asp	AAT Asn	2112
GCT Ala 705	TTA Leu	AGT Ser	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT Tyr	ATA Ile 720	2160
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	2208
гÀя	гÀа	Mec	AAA Lys 740	GIU	AIA	Leu	Glu	745	Gln	Ala	Glu	Ala	Thr 750	Lys	Ala	2256
ile	116	755	TAT Tyr	Gin	Tyr	Asn	Gln 760	Tyr	Thr	Glu	Glu	Glu 765	Lys	Asn	Asn	2304
Ile	770	Phe	AAT Asn	Ile	Asp	775	Leu	Ser	Ser	Lys	Leu 780	Asn	Glu	Ser	Ile	2352
785	riàa	Ата	ATG Met	lle	790	Ile	Asn	Lys	Phe	<b>Leu</b> 795	Asn	Gln	Суѕ	Ser	<b>Val</b> 800	2400
Ser	Tyr	Leu	ATG Met	Asn 805	Ser	Met	Ile	Pro	Tyr 810	Gly	Val	Lys	Arg	Leu 815	Glu	2448
Asp	Phe	Asp	GCT Ala 820	Ser	Leu	Lys	Asp	Ala 825	Leu	Leu	Lys	Tyr	Ile 830	Tyr	Asp	2496
Asn	Arg	61y 835	ACT Thr	Leu	Ile	Gly	Gln 840	Val	yab	Arg	Leu	Lys 845	Asp	Lys	Val	2544
Asn	Asn 850	Thr	CTT	Ser	Thr	Asp 855	Ile	Pro	Phe	Gln	<b>Leu</b> 860	Ser	Lys	Tyr	Val	2592
GAT Asp 865	TAA Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880	2640
CCT Pro	CAA Gln	TCT Ser	AAA Lys	GTT Val 885	AAA Lys	AGA Arg	CAA Gln	ATA Ile	TTT Phe 890	TCA Ser	GGC Gly	TAT Tyr	CAA Gln	TCT Ser 895	GAT Asp	2688
ATT Ile	GAT Asp	ACA Thr	CAT His 900	AAT Asn	AGA Arg	ATT Ile	Lys	GAT Asp 905	GAA Glu	TTA Leu	TGA *					2724

### (2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 908 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 600 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 650

WO 98/07864 PCT/GB97/02273

- 79 -

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
705 710 715 720

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
725 730 735

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
770 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 865 870 875 880

Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp 885 890 895

Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu \* 900 905

# (2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3042 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION:1..3042
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1

48

GTT Val	GAC Asp	ATT Ile	GCC Ala 20	lyr	ATC Ile	Lys	Ile	CCA Pro 25	Asn	GCC Ala	GGC	CAG Gln	ATG Met	Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	Pne	Lys	ATT Ile	CAT His	AAC Asn 40	Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC	144
GAT Asp	ACA Thr 50	hue	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
65	гÀг	GIN	vaı	Pro	70	ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	80	240
Asp	ASN	GIU	гÀЗ	Asp 85	AAC Asn	Tyr	Leu	Lys	Gly 90	Val	Thr	Lys	Leu	Phe 95	Glu	288
Arg	ite	Tyr	Ser 100	Tnr	GAC Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864

ANG TIT ANA GAT ATT GCA AGT ACA CTG AAC ANG GCT AAG TCC ATT GTG  Lys Phe Lys Asp 11e Ala Ser Thr Leu Asn Lys Ala Lys Ser 11e Val  290  GGT ACC ACT GCT TCA TTA CAG TAT ATG ANA AAT GTT TIT AAA GAG AAA  GIY THR THR Ala Ser Leu GIN TYP Met Lys Asn Val Phe Lys Glu Lys 315  TAT CTC CTA TCT GAA GAT ACT CTG GGA AAA TIT TCG GTA GAT AAA TTA  TYP Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325  AAA TIT GAT AAG TTA TAC AAA ATG TTA ACA GAG GAT TAC ACA GAG GAT 194  AAA TIT GAT AAG TTA TAC AAA ATG TTA ACA GAG GAT TAC ACA GAG GAT 195  AAA TIT GAT AAG TTT TAT AAA GTA CTA AACA GAG AAT AAT ATT TOT AAT ASN PHE VAL Lys Phe Phe Lys Val Leu Asn Arg Lys Thr TYP Leu Asn 355  TIT GAT AAA GCC GTA TIT AAA GTA ATA ATA GTA CTA AAA ACA TAT TYP  ACA ATA TAT GAT GAG TTT AAA GTA AAT ATA GTA CTA AAA ACA TAT TYP  ACA ATA TAT GAT GAG TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC THR 11e TYP ASP Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 370  TTT AAT GGT CAA AAT ACA GAA ATT AAT ATA GAA ATA TAT ACA AAA TAC AAA  ATA TAT GAT GAA TACA GAA AAT ACA AAAT TTA CAA AAA ACA  ATA TAT GAT GAA TACA GAA ATT AAT ATA AAT ATA GAA TTTA TAC AAA  ACA ATA TAT GAT GAA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC THR 11e TYP ASP Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 370  TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATA GAA TTT TAC TAAA CTA Phe Asn Gly Gla Asn Thr Glu Ile Asn Asn Ame Aca ATT TTA CT AAA CTA Phe Asn Gly Gla Asn Thr Glu Ile Asn Asn Ame Aca TTT TACT AAA CTA ACA ATA TAT GCG CAA TAT ACA GAA ATT TAT ATA AAA ACT ATA ATA GGT CAA AAA ACA AAA TACA GAA ATT AAT AATA AAA AAT TTA ACT GAA GAA  AAA AAT TTT ACT GCA TTG TTT GAA TTT TAT AAA AAT TTA CAA AAA  AAA TTT GGT CAA AAA ACA GAA AAT AAT AATA AATA	<u>-</u> ,		
TIT CITC CTA TOT GAA GAT ACA TOT GGA AAA TIT TOG GTA GAT AAA TITA CTC CTA TOT GAA GAT ACA TOT GGA AAA TIT TOG GTA GAT AAA TITA CAA AAA TIT ACA AAA TITA CAAA ATA TITA GAT AAA TITA CAAA ATA TAAA CAAA ATA TAAA CAAA ATA TAAA CAAA ATA TAAA CAAAA ATA TAAA CAAAAA ATA TAAAAAAAA	290 295 300	er Ile Val	912
AAA TIT GAT CAA AAT ACA GAA ATT AAT AAT AAT ATT ACT AAA CTA AAT TIT ACT CAA AAT ACT CAT AAA AAT TT TAC CAA AAT TAC CAA AAT TAC AAA ATT CAA AAT TAC AAA ATT CAA CA GAA CAA TAT ACT CAT CAC ACA CAC CAC CAC C	305 310 315 Agn val Phe L	ys Glu Lys 320	960
AAT TIT GIT AGG TIT TIT AAA GIA CIT AAC AGA AAA ACA TAT TIG AAT AAG AAA AACA TAT TIG AAT AACA AAA AACA TAT TIG AAT AACA AAA AACA AAA TAT TAC AACA AAA AA	325 330	Sp Lys Leu 335	1008
TIT GAT AAA GCC GTA TIT AAG ATA AAT ATA GTA CCT AAG GTA AAT TAC Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 375 375 380 375 380 375 380 375 380 380 380 380 400 195 400 19	340 345 345	or Glu Asp 50	1056
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA GCA AAC Thr ile Tyr Asp Gly Phe Ash Leu Arg Ash Thr Ash Leu Ala Ala Ash 385  TTT AAT GGT CAA AAT ACA GAA ATT AAT ATA ATG AAT TTT ACT AAA CTA Phe Ash Gly Gln Ash Thr Glu Ile Ash Ash Met Ash Phe Thr Lys Leu 410  AAA AAT TTT ACT GGA TTG TTT GAA TTT AAT AAT ATG AAT TGT GTA AGA Lys Ash Phe Thr Gly Leu Phe Glu Phe Thr Lys Leu Leu Cys Val Arg 420  GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Ash Lys 445  ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TAT TAT TAC AAA GTT Ile Glu Gly Arg Cys Asp Gly Ala Leu Ash Asp Leu Cys Ile Lys Val 455  AAT AAT TGG GAC TTG TTT TAGT CCT TCA GAA GAT AAT TTT ACT AAT ASh ASh Trp Asp Leu Phe Phe Ser Pro Ser Glu Ash Ash Ash Phe Thr Ash 465  GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA ASP Leu Ash Lys Gly Glu Glu Ile Thr Ser Asp Thr Ash Ile Glu Ala 470  GAT CTA AAT AAA GGA GAA GAA ATT ACT CTA GAT AAT ATA TATA GAA GCA ASP Leu Ash Lys Gly Glu Glu Ile Thr Ser Asp Thr Ash Ile Glu Ala 485  GCA GAA GAA AAT ATT AGT TTA GAT TTA ACA TCT GAT ACT AAT ATA GAA GCA ASP Leu Ash Lys Gly Glu Glu Ile Thr Ser Asp Thr Ash Ile Glu Ala 495  GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC Ala Glu Glu Ash Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505  TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Ash Phe Ash Phe Ash Phe Ash Phe Ash Ash GIU Pro Glu Ash Ile Ser Ile Glu Arg Phe 515  AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT 520  CCT AAT GGA AAA AAG TAT GAG TTA GAT CTA GAT ATA TATA GAA AGA TTT 545  CCT AAT GGA AAA AAG TAT GAG TTA GAT TAT GCT AAT ATA GAA AGA TTT 546  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 546  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 546  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 546  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ACT TTC HIS TYP	355 360 Ash Arg Lys Thr Ty	r Leu Asn	1104
385  390  395  395  396  397  398  398  397  398  398  399  397  398  398	370 375 The Nam Tie Val Pro Lys Va	l Asn Tyr	1152
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA AGA AGG ATA ATA ACT TCT AAA ACT ACT Lys Ser Leu Asp Lys Gly Tyr Asn Lys Asp Leu Cys Val Arg 425  GGG ATA ATA ACT TCT AAA ACT ACT ACT Lys Ser Leu Asp Lys Gly Tyr Asn Lys Add Tyr Asn Lys Add Act	385 390 395	a Ala Asn 400	1200
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AAG GIV Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 440  ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT 1392  ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA GTT 1450  ATT GGIV Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 455  AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT ASN ASN Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 480  GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA 485  GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA 485  GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC 485  GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC 1536  GCA GAA GAA AAT ATT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA ASN Ile Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 510  TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA ASN Leu Ser 520  AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT TCA ASN Ile Glu Asn Leu Ser 530  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TCC CAT TAT 1680  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TCC CAT TAT TCA ATA GAA AGA TTT TCA ASN Ile Glu Arg Phe 7545	405 410	Lys Leu 415	1248
ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TTA ACT AAA GTT 1392  ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT AAT GAT ATA TAT TAT ACT AAA GTT 1392  ASS ASS ASS ASS ASS ASS ASS ASS ASS AS	420 425 Leu Leu Cys	Val Arg	1296
AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT ACT AAT ASN ASN Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 480  GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA ASP Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 495  GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC 1536  GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC 1536  TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA 1584  Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 525  AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT 1632  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 1680  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 1680  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 1680  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 1680	435 440 Let Asp Lys Gly Tyr	Asn Lys	1344
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GCA 1488  GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC 1536  GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC 1536  TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA 1584  Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520  AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530  CCT AAT GGA AAA AAG TAT GAG TTA GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr	450 455 Lett Ash Asp Lett Cys Ile	Lys Val	1392
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACC 1536  TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA 1584  Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 525  AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 535  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 1680  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 1680  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT 1680	465 470 SET GIU ASP ASN Phe	Thr Asn 480	1440
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TCA Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser S15  AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe S30  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr  1680	485 490 Asp Thr Asn Ile	Glu Ala 495	1488
AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TTT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr  1680	500 505 510	Leu Thr	1536
530  535  536  537  538  540  CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr  545	515 520 Ser He Glu Asn 525	Leu Ser	1584
545 S50 Star Asp Lys Tyr Thr Met Phe His Tyr	530 535 S40	Arg Phe	1632
	545 Fig. 12 See Asp Lys Tyr Thr Met Phe I	His Tyr	1680

CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	Pne	GAA Glu	CAT	GGT Gly	AAA Lys 570	Ser	AGG Arg	ATT	GCT Ala	TTA Leu 575	ACA Thr	1728
AAT Asn	TCT	GTT Val	AAC Asn 500	GAA Glu	GCA Ala	TTA Leu	TTA Leu	AAT Asn 585	Pro	AGT Ser	CGT	GTT Val	TAT Tyr 590	ACA Thr	TTT Phe	1776
TTT Phe	TCT Ser	TCA Ser 595	GAC Asp	TAT Tyr	GTA Val	AAG Lys	AAA Lys 600	GTT Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 605	GAG Glu	GCA Ala	GCT Ala	1824
Met	Phe 610	TTA Leu	GIY	Trp	Val	Glu 615	Gln	Leu	Val	Tyr	Asp 620	Phe	Thr	Asp	Glu	1872
ACT Thr 625	AGC Ser	GAA Glu	GTA Val	AGT Ser	ACT Thr 630	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 640	1920
ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 645	CCT	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	AAA Lys	1968
GAT Asp	GAT Asp	TTT Phe	GTA Val 660	GGT Gly	GCT Ala	TTA Leu	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 670	CTG Leu	TTA Leu	2016
GAA Glu	TTT Phe	ATA Ile 675	CCA Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT Leu	2064
GTA Val	TCA Ser 690	TAT Tyr	ATT Ile	GCG Ala	AAT asa	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	ACA Thr	ATA Ile	GAT Asp	AAT Asn	2112
GCT Ala 705	TTA Leu	AGT Ser	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT Tyr	ATA Ile 720	2160
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	ÀAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	2208
AAA Lys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala	2256
ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	TAA Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	TAA Asn	AAT Asn	2304
ATT Ile	AAT Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile	2352
AAT Asn 785	AAA Lys	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800	2400
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu	2448
GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GAT Asp	2496

		83	5			- Gry	840	Val	. Asp	Arg	, Leu	Lys 845	Asp	Ly	A GTT S Val	2544
AA' As:	T AAT n Asr 0 850		A CTI	r AG7 1 Se1	T ACA	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	Ser	Lys	TAC Ty	GTA Val	2592
GA: Asp 865	TAAT Asn	CAA 1 Glr	A AGA 1 Arg	TTA Leu	TTA Leu 870	Jer	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT	ATT Ile	AAG Lys	TCA Ser	GGC Gly 880	2640
				885	ALG	vra	uis	TYE	890	Gin	His	Asp	Glu	Ala 895		2688
GAC Asp	AAC Asn	AAA Lys	Phe 900	U3!!	AAA Lys	GAA Glu	CAA Gln	CAA Gln 905	AAC Asn	GCG Ala	TTC Phe	TAT Tyr	GAG Glu 910	ATC Ile	TTA Leu	2736
CAT His	TTA Leu	CCT Pro 915	AAC Asn	TTA Leu	AAC Asn	GAA Glu	GAA Glu 920	CAA Gln	CGA Arg	AAC Asn	GCC Ala	TTC Phe 925	ATC Ile	CAA Gln	AGT Ser	2784
TTA Leu	AAA Lys 930	GAT Asp	GAC Asp	CCA Pro	AGC Ser	CAA Gln 935	AGC Ser	GCT Ala	AAC Asn	CTT Leu	TTA Leu 940	GCA Ala	GAA Glu	GCT Ala	AAA Lys	2832
AAG Lys 945	CTA Leu	AAT Asn	GAT Asp	GCT Ala	CAG Gln 950	GCG Ala	CCG Pro	AAA Lys	GTA Val	GAC Asp 955	AAC Asn	AAA Lys	TTC Phe	AAC Asn	AAA Lys 960	2880
GAA Glu	CAA Gln	CAA Gln	AAC Asn	GCG Ala 965	TTC Phe	TAT Tyr	GAG . Glu	тте	TTA Leu 970	CAT His	TTA Leu	CCT . Pro .	Asn	TTA Leu 975	AAC Asn	2928
GAA Glu	GAA Glu	CAA Gln	CGA Arg 980	AAC Asn	GCC Ala	TTC . Phe	TIG (	CAA Gln 985	AGT Ser	TTA . Leu	AAA (	Asp A	GAC Asp 990	CCA Pro	AGC Ser	2976
CAA Gln	AGC Ser	GCT Ala 995	AAC Asn	CTT Leu	TTA ( Leu /	GCA ( Ala (	GAA ( Glu ) LOOO	GCT :	AAA . Lys :	AAG ( Lys )	Leu A	AAT ( Asn /	GAT ( Asp )	GCT Ala	CAG Gln	3024
Ala	CCG Pro 1010	AAA Lys	GTA ( Val .	GAC Asp	TAG											3042

# (2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1014 amino acids
    (B) TYPE: amino acid
    (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glú Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435

Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480

Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495

Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505

Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515

Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 540

Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555

Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 575

Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590

Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640

Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
645 650 655

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
705 710 715 720

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
725 730 735

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
740 745 750

- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
  770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 870 875 880
- Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895
- Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu 900 905 910
- His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925
- Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935 940
- Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955 960
- Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 965 970 975
- Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990
- Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005
- Ala Pro Lys Val Asp \* 1010
- (2) INFORMATION FOR SEQ ID NO: 19:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 3509 base pairs
    - (B) TYPE: nucleic acid
    - (C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)
  - (ix) FEATURE:
    - (A) NAME/KEY: CDS
    - (B) LOCATION:1..3509
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

	1				ATA Ile 5	7101	. ver	1 FIN	= AS	n 1y 1	.0	sn A	rsb	Pro	Il	e A	sp 15	Asn		48
				20	ATG Met		910		2	o Pn 5	e Al	La A	rg (	Gly	Th:	r G.	ly	Arg		96
	_		35		TTT Phe	-, 0		40	. As	p Ar	g 11	e T	rp ]	11e 45	Ile	Pı	0	Glu		144
		50			GGA Gly	- , -	55	PLO	GIL	ı As	p Ph	e As	sn I 50	ys	Ser	Se	r	Gly		192
6	5				GAT (	70	Cys	GIU	TYE	туз	7 AS	p Pı S	co A	sp	Tyr	Le	u i	Asn 80		240
				.,	AAG 1 Lys 1 85		*16	FILE	rea	90	Thi	r Me	t I	le	Lys	Le	ս 1 5	?he		288
		J -	1	00	CA A	.y.S	-10	Deu	105	GIU	Lys	: Le	u L	eu (	Glu 110	Me	t I	le		336
		11	īś -		CT 1	<i>y</i> • ·	Deu	120	Авр	Arg	Arg	y Va	1 P:	ro 1 25	Leu	Gli	ı G	lu		384
	13	0			TT G le A		135	val	inr	val	Asn	Ly:	s Le O	u 1	lle	Ser	A	sn		432
145				0.		50	Jys 1	Lys	GIÀ	TIE	Phe 155	Ala	a As	n L	eu	Ile	1	le 50		480
	2			16	_	•	icu ,	, 11C	GIU	170	GIU	Thr	· Il	e A	sp	Ile 175	G)	ly		528
			18	ō	MT GO	.a J	CI N	1	185	GIÀ	Pne	Gly	Gl	y I	le i 90	Met	G1	.n		576
		19	5		A GA	<b>.</b>	2	00	er .	vaı	Pne	Asn	209	ı Va	al (	Sln	Gl	u		624
	210				T AT	2:	15		ug /	arg (	GTÀ	1yr 220	Phe	: Se	er A	rab	Pr	0		672
225				- 120		0	Lu Di	eu 1	Te :	ils	235	Leu	His	Gl	y L	eu	Ty:	r O		720
GGC		-1-		245	5	ש איני	u Pi	20 1	1e v	250	Pro J	Asn	Glu	Ly	s L 2	ys 55	Phe	2		768
TTT .	ATG Met	CAA Gln	Ser 260		A GAT	GC Al	T AT	. C G	AG G ln A 65	CA C	SAA ( Slu (	GAA Glu	CTA Leu	TA Ty 27	r T	CA hr	TT: Phe	<u> </u>	8	316

GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
TAT Tyr	GAT Asp 290	Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
						TTA Leu										1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632

54	5		10 .	Deu	ASI	55	e ar	g As	р 11	A AG e Se	r Le 55	u Th 5	ır Se	er S	er F	he	Asp 560	1680
715	, P	.a D	cu i	JE U	565	5	L AS	и пА	s va	T TA 1 Ty: 57	r Se O	r Ph	le Ph	e Se	er M 5	et 75	Asp	1728
-7		C 13,	, 5	80	nia	. Yai	ı Lys	s va.	58!	_	ı Ala	a Gl	y Le	u Ph	e A	la	Gly	1776
••	p •u	59	5	7111	116	val	. ASI	600	)	r GT/ e Val	l Ile	e Gl	u Al 60	a As 5	n L	ys	Ser	1824
,,,,	61	0	.с л	чэр	nys	110	615	ASP	) 116	A TCT	Leu	62	e Va O	l Pr	o Ty	r	Ile	1872
629	5	- 774	<b>u</b>	cu	W211	630	GIÀ	ASI	GIU	ACA Thr	635	Lys	s Gl	y As:	n Ph	ie	Glu 640	1920
75.			.e	Iu	645	wia	GIĀ	ATA	Ser	ATT Ile 650	Leu	Let	ı Glı	ı Phe	e Il 65	e . 5	Pro	1968
GI	. Dec	ı ne	6	60	PIO	vai	var	GTÅ	665		Leu	Leu	Glu	Se) 670	Ту	r :	Ile	2016
	11011	67	5	344 1	uys	116	116	680	inr	ATA Ile	Asp	Asn	Ala 685	Lev	Th	r I	ŗÀ2	2064
	690	<b>U</b> 1	ים ב	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	rīþ	Set	695	met	Tyr	GGA Gly	Leu	11e 700	Val	Ala	Gli	n I	,rb	2112
705	502			-	7311	710	GIN	Pne	Tyr	ACA Thr	715	Lys	Glu	Gly	Met	: T	yr 20	2160
Dy S	710	nec	i ns	7	25	GIN	AIA	GIN	Ala	TTG Leu 730	Glu	Glu	Ile	Ile	Lys 735	T	yr	2208
Arg	TYL	WDI.	74	0	yr :	ser	GIU	rys	745	AAG Lys	Ser	Asn	Ile	<b>Asn</b> 750	Ile	A	sp	2256
1110	veir	7 <b>5</b> 5	11	C A	9N 3	ber	rys	760	Asn	GAG Glu	Gly	Ile	<b>Asn</b> 765	Gln	Ala	I.	le	2304
rap	770	116	AS	u A	sn e	ne	775	Asn	GIA	TGT Cys	Ser	Val 780	Ser	Tyr	Leu	Me	et	2352
785	nys	Mec	116	e P	7	90	Ala V	/al (	Glu :		<b>Leu</b> :	Leu	Asp	Phe	Asp	A:	sn O	2400
ACT Thr	CTC Leu	AAA Lys	AA/ Lys	3 A:	AT T sn L 05	TG : leu l	rta ; Leu ;	AT '	ryr .	ATA ( Ile ; 810	GAT ( Asp (	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	T? Ty	AT 'T	2448

TTC Lev	ATI	GG/	A AG2 / Se1 820	. WIG	A GAA A Glu	TAT Tyr	GAA Glu	AAA Lys 825	Ser	AAA Lys	GT#	A AA? L As:	LAA 1 Lys 830	Ту	TTG Leu	2496
AAA Lys	ACC	ATT Ile 835	. 1.10	CCC Pro	TTI Phe	GAT Asp	CTT Leu 840	Sei	ATA	TAT	ACC Thr	AA1 Asr 845	Asp	AC#	ATA Ile	2544
CTA Leu	Ile 850	GIU	ATC Met	TTI Phe	AAT Asn	AAA Lys 855	TYT	AAT Asn	AGC Ser	GAA Glu	ATT Ile 860	: Leu	AAT Asn	AAT Asn	ATT	2592
ATC Ile 865	TTA Leu	AAT Asn	TTA Leu	AGA Arg	TAT Tyr 870	ьys	GAT Asp	AAT Asn	AAT Asn	TTA Leu 875	ATA Ile	GAT Asp	TTA Leu	TCA Ser	GGA Gly 880	2640
TAT Tyr	GGG Gly	GCA Ala	AAG Lys	GTA Val 885	GIU	GTA Val	TAT Tyr	GAT Asp	GGA Gly 890	GTC Val	GAG Glu	CTT	AAT Asn	GAT Asp 895	AAA Lys	2688
AAT Asn	CAA Gln	TTT Phe	AAA Lys 900	TTA Leu	ACT Thr	AGT Ser	TCA Ser	GCA Ala 905	AAT Asn	AGT Ser	AAG Lys	ATT Ile	AGA Arg 910	GTG Val	ACT Thr	2736
CAA Gln	AAT Asn	CAG Gln 915	AAT Asn	ATC Ile	ATA Ile	TTT Phe	AAT Asn 920	AGT Ser	GTG Val	TTC Phe	CTT Lėu	GAT Asp 925	TTT Phe	AGC Ser	GTT Val	2784
AGC Ser	TTT Phe 930	TGG Trp	ATA Ile	AGA Arg	ATA Ile	CCT Pro 935	AAA Lys	TAT Tyr	AAG Lys	AAT Asn	GAT Asp 940	GGT Gly	ATA Ile	CAA Gln	AAT Asn	2832
TAT Tyr 945	ATT Ile	CAT His	AAT Asn	GAA Glu	TAT Tyr 950	ACA Thr	ATA Ile	ATT Ile	TAA	TGT Cys 955	ATG Met	AAA Lys	AAT Asn	AAT Asn	TCG Ser 960	2880
GGC Gly	TGG Trp	AAA Lys	ATA Ile	TCT Ser 965	ATT Ile	AGG Arg	GGT Gly	AAT Asn	AGG Arg 970	ATA Ile	ATA Ile	TGG Trp	ACT Thr	TTA Leu 975	ATT Ile	2 <b>92</b> 8
GAT Asp	ATA Ile	AAT Asn	GGA Gly 980	AAA Lys	ACC Thr	AAA Lys	Ser	GTA Val 985	TTT Phe	TTT Phe	GAA Glu	TAT Tyr	AAC Asn 990	ATA Ile	AGA Arg	2976
GIU	мsр	995	ser	GIU	TAT Tyr	TTE	1000	Arg	Trp	Phe	Phe	Val 1005	Thr,	Ile	Thr	3024
ASn	AAT Asn 1010	Leu	AAT Asn	AAC Asn	GCT Ala	AAA Lys 1015	ATT	TAT Tyr	ATT Ile	Asn	GGT Gly 1020	Lys	CTA Leu	GAA Glu	TCA Ser	3072
AAT Asn 1025	Inr	GAT Asp	ATT Ile	rys	GAT Asp 1030	TIG .	AGA ( Arg (	GAA Glu	Val	ATT Ile 1035	Ala	AAT Asn	GGT Gly	GAA Glu	ATA Ile 1040	3120
ATA Ile	TTT Phe	AAA Lys	rea	GAT Asp 1045	GGT (	GAT Asp	ATA (	Asp .	AGA Arg 1050	ACA Thr	CAA Gln	TTT Phe	Ile	TGG Trp 1055	Met	3168
AAA Lys	TAT	Pne	AGT Ser 1060	He	TTT / Phe /	AAT / Asn /	Thr (	GAA Glu 1065	TTA . Leu :	AGT (	CAA Gln	Ser	AAT Asn 1070	ATT Ile	GAA Glu	3216
GAA . Glu .	Arg	TAT Tyr 1075	Lys	ATT Ile	CAA '	Ser '	TAT I	AGC Ser	GAA ' Glu '	TAT '	Leu	<b>AAA</b> Lys 1085	GAT Asp	TTT Phe	TGG Trp	3264

GGA Gly	AAT Asn 109		TTA Leu	ATG Met	TAC	AAT Asn 109	ьys	GAA Glu	TAT	TAT Tyr	ATO Met	Phe	AAT Asn	GCG Ala	GGG Gly	3312
110	5	7.011	501	* 7 *	1110	) TÀR	Leu	rys	Lys	Asp 111!	Ser 5	CCT	Val	Gly	Glu 1120	3360
	200		AT 9	1125	5 5	TYE	ASI	GIN	Asn 113	Ser	Lys	TAT Tyr	Ile	Asn 1135	Tyr	3408
9			1140	)	GIY	GIU	гÀа	1145	ile	Ile	Arg	AGA Arg	Lys 1150	Ser	Asn	3456
TCT Ser	CAA Gln	TCT Ser 1155		AAT Asn	GAT Asp	usb	ATA Ile 1160	vaı	AGA Arg	AAA Lys	GAA Glu	GAT Asp 1165	Tyr	ATA Ile	TAT Tyr	3504
CTA Leu	GA															3509

### (2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1169 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 200 Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 315 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 470 Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 505 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530 535

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 560

Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575

Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585 590

Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605

Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 620

Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 625 635 640

Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655

Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670

Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685

Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700

Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720

Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 735

Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp
740 745 750

Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile 755

Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 780

Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795 800

Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815

Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830

Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845

Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile 850 855 860

Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 865 870 880 Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895

Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 910

Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925

Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940

Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960

Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975

Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990

Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005

Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020

Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040

Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1045 1050 1055

Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070

Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085

Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1090 1095 1100

Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120

Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 1135

Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150

Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165

Leu

#### (2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2574 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

- 95 -

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2574

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

***	1	.0 •			5	m AS	n Ph	e As	10	r Asi	n As	p Pr	o Il	e As 1	_	48
A# As	AT AA	T A	re r	TT AT le Me 20	rg Al et Me	G GAC	G CC	T CC	o Pue	GCC Ala	G AG	A GG g Gl	T AC	r Gl	G AGA y Arg	96
TA Ty	T TA		AA G /s A 85	CT TT la Ph	TT AA ne Ly	A ATO	C ACI	. wai	CGI Arg	ATI Ile	TGC Tr	3 AT	e Ile	A CC	G GAA D Glu	144
AG Ar	A TA g Ty 5		T T	MT GO	A TA y Ty	T AAA r Lys 55	PIC	GAC Glu	GAT Asp	TTI Phe	AAT Asn 60	Lys	A AGT	TC(	GGT Gly	192 ·
AT Il 6		T AA e As	T AC	SA GA ng As	T GT p Va. 7	TGT l Cys 0	GAA Glu	TAT Tyr	TAT Tyr	GAT Asp 75	Pro	GAT Asp	TAC Tyr	TTA Leu	TAA Anna ana ana ana ana ana ana ana ana	240
AC'	r Aa' r Asi	r GA n As	T AF	A AA 's Ly 8	3 W21	r ATA n Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	: AAG : Lys	Leu 95	Phe	288
AA: Asi	AG/	A AT	C AA e Ly 10	6 3e	A AA/ r Lys	A CCA B Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT	336
ATA Ile	AAI : Asr	GG Gl; 11	,	A CC	TA1	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	384
TTT	AAC Asn 130		A AA r As	C AT	C GCT	Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn	432
CCA Pro 145	017	GA:	A GTO	G GAC	G CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160	480
TTT Phe	GGA Gly	Pro	GG(	CCA Pro 165	ATT	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly	528
ATA Ile	CAA Gln	AAT Asn	CAT His	FILE	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GCC	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln	576
ATG Met	AAG Lys	TTT Phe 195	Cya	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA '	TTT Phe	Asn .	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624
AAC Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 2	AAT . Asn .	AGA Arg	CGT ( Arg (	Gly '	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro	672
GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA ( Glu )	CTT : Leu	ATA (	HIS A	GTT : Val 1 235	TTA ( Leu )	CAT His	GGA Gly	Leu	TAT Tyr 240	720

GGC Gly	ATI	AAA Lys	GTA Val	GAT Asp 245	) Asp	TTA Leu	CCA Pro	ATI	GTA Val 250	Pro	AAI Asn	GAA Glu	AAA Lys	Lys 255	TTT Phe	768
Pne	Mec	. GIN	260	ing	Авр	Ala	lle	G1n 265	Ala	Glu	Glu	Leu	Tyr 270	Thr	TTT Phe	816
GIY	GIÀ	275	Asp	Pro	ser	IIe	11e 280	Thr	Pro	Ser	Thr	Asp 285	Lys	Ser	ATC	864
TAT Tyr	GAT Asp 290	гÀг	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	91,2
AAG Lys 305	val	TTA Leu	GTT Val	TGC	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
inr	370	Ala	TCT Ser	туг	Phe	Ser 375	Asp	Ser	Leu	Pro	Pro 380	Val	Lys	Ile	Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	GIn	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536

GA? Ası	r TTT	AA? Asi 519	ı va.	A GAT L Asi	r GT	CCA L Pro	A GTZ Va: 520	⊥ Туз	r GAJ	A AAI 1 Ly:	A CA	A CCC n Pro 525	o Al	T AT a Il	A AAA e Lys	1584
AAA Lys	A ATT	Phe	ACA Tha	A GAI	GAZ Glu	AAT Asn 535	Thi	C ATO	TTT Phe	CAJ Gli	A TA: 1 Ty: 540	r Lei	A TA	C TC r Se	T CAG r Gln	1632
ACA Thr 545	Pne	CCI Pro	CTA Leu	GAT Asp	TATA 11e 550	Arg	GAT Asp	T ATA	AGT Ser	TTA Lev 559	Thi	A TCT	TC:	A TT	GAT Asp 560	1680
GAT Asp	GCA Ala	TTA Leu	TTA Leu	TTI Phe 565	Ser	AAC Asn	AAA Lys	GTI Val	TAT Tyr 570	Sex	TTT Phe	TTT Phe	TC: Sei	T ATO	G GAT Asp	1728
TAT Tyr	ATT Ile	AAA Lys	ACT Thr 580	Ala	AAT Asn	AAA Lys	GTG Val	GTA Val 585	Glu	GCA Ala	GGA Gly	TTA Leu	TT7 Phe 590	ala e	A GGT A Gly	1776
TGG Trp	GTG Val	AAA Lys 595	CAG Gln	ATA Ile	GTA Val	AAT Asn	GAT Asp 600	Phe	GTA Val	ATC Ile	GAA Glu	GCT Ala 605	Asr	Lys	AGC Ser	1824
AAT Asn	ACT Thr 610	Met	GAT Asp	AAA Lys	ATT	GCA Ala 615	GAT Asp	ATA Ile	TCT Ser	CTA Leu	ATT Ile 620	Val	CCT	TAT	ATA Ile	1872
GGA Gly 625	Leu	GCT Ala	TTA Leu	AAT Asn	GTA Val 630	GGA Gly	AAT Asn	GAA Glu	ACA Thr	GCT Ala 635	AAA Lys	GGA Gly	AAT Asn	TTI Phe	GAA Glu 640	1920
AAT Asn	GCT Ala	TTT Phe	GAG Glu	ATT Ile 645	GCA Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	CTA Leu	CTA Leu	GAA Glu	TTT Phe	ATA Ile 655	CCA Pro	1968
GAA Glu	CTT Leu	TTA Leu	ATA Ile 660	CCT Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	TTT Phe	TTA Leu	TTA Leu	GAA Glu	TCA Ser 670	TAT Tyr	ATT Ile	2016
GAC Asp	AAT Asn	AAA Lys 675	AAT Asn	AAA Lys	ATT Ile	ATT Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	TTA Leu	ACT Thr	AAA Lys	2064
AGA Arg	AAT Asn 690	GAA Glu	AAA Lys	TGG Trp	AGT Ser	GAT Asp 695	ATG Met	TAC Tyr	GGA Gly	TTA Leu	ATA Ile 700	GTA Val	GCG Ala	CAA Gln	TGG Trp	2112
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720	2160
AAG Lys	GCT Ala	TTA Leu	TAA Rsn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr	2208
AGA Arg	TAT Tyr	AAT Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA Glu	AAA Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	Ile	AAC Asn 750	ATC Ile	GAT Asp	2256
TTT Phe	Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	Lys	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	ATT Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile	2304
GAT Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	Phe	ATA / Ile / 775	AAT Asn	GGA Gly	TGT Cys	TCT Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met	2352

AAA Lys 785	TA2 TY3	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA Leu	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448
TTG	ATT Ile	GGA Gly	AGT Ser 820	GCA Ala	GAA Glu	TAT Tyr	GAA Glu	AAA Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC Tyr	TTG Leu	2496
AAA Lys	ACC Thr	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
CTA Leu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	AAT Asn	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser							2574

- (2) INFORMATION FOR SEQ ID NO: 22:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 858 amino acids
    - (B) TYPE: amino acid
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: protein
  - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu
35 40 45

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly
50 55 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185 190

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220

Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 235 240

Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255

Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 260 270

Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285

Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300

Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 310 315 320

Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325

Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350

Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365

Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 380

Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390 395 400

Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415

Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430

Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435

Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455

Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480

Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495

Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510

Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530

- 100 -

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 550 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 600 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser

#### (2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1644 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

WO 98/07864 PCT/GB97/02273

- 101 -

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..1644

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

										-	:					
146	1	o va	ı ini	2 116	ası S	l AST	ı Pne	e Asr	1 Tyr	Asr	ı Asp	Pro	o Ile	2 As <sub>1</sub>	_	48
AA: Asi	T AA:	r AT	T ATT	: Met	ATG Met	GAG Glu	CCT Pro	CCA Pro 25	Phe	GCC Ala	AGA Arg	GGT Gly	Thi	Gly	G AGA / Arg	96
TA]	TAT	AA Lys 35	S Ala	TTI Phe	AAA Lys	ATC	ACA Thr	Asp	CGT Arg	ATI	TGG	ATA Ile 45	Ile	CCC Pro	G GAA	144
AGA Arg	TAT Tyr 50	1111	TTT Phe	GGA Gly	TAT	AAA Lys 55	CCI	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	Lys	AGT Ser	TCC Ser	GGT Gly	192
ATT Ile 65	Pne	AA1	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT	TAT Tyr	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80	240
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95	TTT	288
ASII	Arg	116	Lys 100	ser	гåа	PTO	Leu	105	Glu	Lys	Leu	Leu	Glu 110	Met	Ile	336
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	384
TTT Phe	AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn	432
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys .	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160	480
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly	528
ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	Gly	ATA Ile 190	ATG Met	CAA Gln	576
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA ' Glu '	Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624
Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT . Ser	ite i	TTT Phe 2	AAT Asn	AGA Arg	CGT   Arg	Gly	TAT Tyr 220	TTT ' Phe	TCA Ser	GAT Asp	CCA Pro	672

GCC Ala 225	Den	ATA Ile	. TTA Leu	AIG Met	CAT His 230	GIU	CTT Leu	ATA Ile	CAT His	GT1 Val 235	Leu	CAT His	GG#	TTA Leu	TAT Tyr 240	720
GGC Gly	ATT	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT	768
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	Ala	G <b>AA</b> Glu	GAA Glu	CTA Leu	TAT Tyr 270	Thr	TTT	816
GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
TAT	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
TAA neA	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	Asn .	GAC Asp 485	TTC Phe	CCT Pro	ATA . Ile .	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488

- 103 -

TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GIU	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632
		CCT Pro														1644

### (2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 548 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
180 185 190

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 520 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 535 Thr Phe Pro Leu 545

- 105 -

## (2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2616 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

#### (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..2616

# (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

ATG Met	011	TTO Phe	GTC Val	G AAC l Asr	AA( Lys	G CAC	TTC Phe	AA(	TAT	Lys	GAC Asp	C CC1	GT#	AAA Ası	GGT Gly	4	8 .
GTT Val	GAC Asp	ATT Ile	GCC Ala 20	. ryı	: ATC	Lys	ATI	CCA Pro	Asn	GCC Ala	GGC Gly	CAC Glr	ATG Met	Glr	CCG Pro	9	6
val	Бys	35	PILE	: TÀ2	116	nis	40	Lys	Ile	Trp	Val	1le 45	Pro	Glu	CGC Arg	14	4
7.02	50		1114	nsii	FIU	55	Gru	Gly	Asp	Leu	Asn 60	Pro	Pro	Pro	GAA Glu	19:	2
65	-,-	0111	vul	110	70	361	TYE	ıyr	Asp	75	Thr	Tyr	Leu	Ser	80	24(	)
		0.10	Lys	GAT Asp 85	ASII	IYI	Leu	гÀа	90 GIA	Val	Thr	Lys	Leu	Phe 95	Glu	288	3
arg	116	-y-	100	ACT Thr	Asp	Leu	GIY	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val	336	i
9	<b>-</b> 1	115	110	TTT Phe	110	GIY	120	ser	Tnr	Ile	Asp	Thr 125	Glu	Leu	Lys	384	
Val	130	vaħ	1111	AAC Asn	Cys	135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Tyr	432	
145	361	GIU	Gru	CTT Leu	150	ren	vai	lie.	Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160	480	
ATC (	3211		GIU	165	nys	ser	rne	GIÀ	H18 170	Glu	Val	Leu	Asn	Leu 175	Thr	528	
CGT A	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	GIN	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	Ser	CCA Pro .	GAC Asp	TTC Phe	576	

ACC Thr	TTC Phe	GGT Gly 195	FILE	GAC Glu	G GAG	AGC Ser	CTC Leu 200	I GIU	GTI Val	GAT Asp	ACC Thr	AAC Asn 205	Pro	CTC Lev	TTG Leu	624
G <b>GT</b> Gly	GCA Ala 210	GIY	Lys	TTC Phe	GCA Ala	ACT Thr 215	' Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	Leu	GCA Ala	CAC His	GAG Glu	672
225	116	nis	Ala	GIY	230	Arg	Leu	Tyr	Gly	1le 235	Ala	Ile	Asn	Pro	AAC Asn 240	720
Arg	Val	Pne	гÀг	245	ASN	Tnr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	TTA Leu	768
GIU	vai	ser	260	GIU	GIU	Leu	Arg	Thr 265	Phe	Gly	Gly	CAT His	Asp 270	Ala	Lys	816
PHE	116	275	ser	Leu	GIN	GIU	Asn 280	Glu	Phe	Arg	Leu	TAC Tyr 285	Tyr	Tyr	Asn	864
гуз	290	ьys	Asp	116	Ala	295	Thr	Leu	Asn	Lys	Ala 300	AAG Lys	Ser	Ile	Val	912
305	Inr	inr	Ala	ser	310	GIn	Tyr	Met	Lys	Asn 315	Val	TTT Phe	Lys	Glu	Lys 320	960
Tyr	Leu	Leu	Ser	325	Asp	Thr	Ser	Gly	Lys 330	Phe	Ser	GTA Val	Asp	Lys 335	Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
Asn	Pne	355	Lys	Phe	Phe	Lys	Val 360	Leu	Asn	Arg	Lys	ACA Thr 365	Tyr	Leu	Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
A <b>A</b> A Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	T Y Y Y Y Y	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	TAA Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392

465	,			GAT Asp	470	£116	1111	ASII	Asp	475	ASI	n Lys	s G]	ly G	lu	Glu 480	1440
		-		ACT Thr 485	7.512	116	GIU	ALG	490	GIu	Gli	ı Asr	ı Il	e S	er 95	Leu	1488
			500	CAA Gln	-71	LYL	Dea	505	Pne	Asn	Phe	: Asp	As 51	n G. 0	lu	Pro	1536
		515		ATA Ile	010	Vall	520	ser	ser	Asp	Ile	1le 525	Gl	y GI	ln	Leu	1584
	530			AAT Asn		535	nig i	PHE	PIO	Asn	Gly 540	Lys	Ly	<b>з Т</b> у	r (	Glu	1632
545		-,-	-1-	ACT Thr	550	rne r	115	ıyr	Leu	555	Ala	Gln	Gli	ı Ph	e (	31u 560	1680
	4	-,,	501	AGG Arg 565	116 /	11a 1	ed 1	nr	570	Ser	Val	Asn	Glu	1 Al 57	a I 5	eu	1728
			580	CGT ( Arg	var .	.yr 1	5	85	rne	ser	Ser	Asp	Tyr 590	Va.	l L	ys	1776
-2-		595	-,	GCT A		6	00 00	ıa ı	me c	rne :	Leu	Gly 605	Trp	Va:	LG	lu	1824
	610		- ,	.ap	6	15	sp G.	IU .	inr :	ser (	31u 520	Val	Ser	Thr	T	hr	1872
GAT A Asp I 625	-,			6	30	11L I.	re I.	re 1	.1e i	35	Cyr :	Ile (	Gly	Pro	64	la 10	1920
TTA A Leu A			6	45	- L	-u 1)	'I LLY	/B A	50 50	rab E	he v	Val (	Gly	Ala 655	Le	eu	1968
ATA T	TT The S		GA G ly A 60	CT G la V	TT A: al II	rr Ci le Le	G TI u Le 66	uG	AA I lu P	TT A	TA C	Pro C	GAG Glu 570	ATT Ile	GC Al	A .a	2016
ATA C Ile P	:	TA T al L 75	TA G eu G	GT A	or T	TT GC ne Al 68	a ne	T G	TA T al S	CA T	yr I	ATT G	CG lla	AAT Asn	AA Ly	.G s	2064
GTT C' Val L	TA A eu T 90	CC G hr V	TT C	AA AG ln Tì	CA AT ar Il 69	C AS	T AA P As	T G	CT T la L	eu S	GT A er L 00	AA A ys A	GA .rg	AAT Asn	GA G1	A u	2112
AAA To Lys T: 705	GG G. rp A	AT G	AG G: lu Va	rc TA al Ty 71	L Dy	A TA	T AT	A G: e Va	al T	CA A hr A	AT T sn T	GG T	TA (	GCA Ala	AA Ly 72	5	2160
GTT AM	AT A	CA CI hr Gi	AG A1 In II	re wa	T CT	A ATA	A AGI	A AJ 9 Ly 73	's Ly	AA A: /s Me	rg A	AA G ys G	lu /	GCT Ala 735	TT	A 1	2208

- 108 -

GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile	2352
AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
					ATT Ile 870		TAA *									2616

#### (2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 872 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 235 230 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445

Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 455 460

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 470 Ile Thr Ser Asp Thr Asm Ile Glu Ala Ala Glu Glu Asm Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 550 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 585 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 650 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 680 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810

WO 98/07864 PCT/GB97/02273

- 111 -

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860

Thr Phe Thr Glu Tyr Ile Lys 865 870

- (2) INFORMATION FOR SEQ ID NO: 27:
  - (i) SEQUENCE CHARACTERISTICS:
    - (A) LENGTH: 2574 base pairs
    - (B) TYPE: nucleic acid(C) STRANDEDNESS: double
    - (D) TOPOLOGY: linear
  - (ii) MOLECULE TYPE: DNA (genomic)

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ATGCCGGTTA CCATCAACAA CTTCAACTAC AACGACCCGA TCGACAACAA CAACATCATC 60 ATGATGGAAC CGCCGTTCGC ACGTGGTACC GGTCGTTACT ACAAGGCTTT CAAGATCACC 120 GACCGTATCT GGATCATCCC GGAACGTTAC ACCTTCGGTT ACAAACCTGA GGACTTCAAC 180 AAGAGTAGCG GGATTTTCAA TCGTGACGTC TGCGAGTACT ATGATCCAGA TTATCTGAAT 240 ACCAACGATA AGAAGAACAT ATTCCTTCAG ACTATGATCA AGTTATTTAA TAGAATCAAA 300 TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA 360 GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA 420 TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA 480 TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT 540 TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA 600 AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT 660 TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTTATAC ATGTTTTACA TGGATTATAT 720 GGCATTAAAG TAGATGATTT ACCAATTGTA CCAAATGAAA AAAAATTTTT TATGCAATCT 780 ACAGATGCTA TACAGGCAGA AGAACTATAT ACATTTGGAG GACAAGATCC CAGCATCATA 840 ACTCCTTCTA CGGATAAAAG TATCTATGAT AAAGTTTTGC AAAATTTTAG AGGGATAGTT 900 GATAGACTTA ACAAGGTTTT AGTTTGCATA TCAGATCCTA ACATTAATAT TAATATATAT 960 AAAAATAAAT TTAAAGATAA ATATAAATTC GTTGAAGATT CTGAGGGAAA ATATAGTATA 1020 GATGTAGAAA GTTTTGATAA ATTATATAAA AGCTTAATGT TTGGTTTTAC AGAAACTAAT 1080 ATAGCAGAAA ATTATAAAAT AAAAACTAGA GCTTCTTATT TTAGTGATTC CTTACCACCA 1140 GTAAAAATAA AAAATTTATT AGATAATGAA ATCTATACTA TAGAGGAAGG GTTTAATATA 1200

TCTGATAAAG	ATATGGAAAA	AGAATATAGA	GGTCAGAATA	AAGCTATAAA	TAAACAAGCT	1260
TATGAAGAAA	TTAGCAAGGA	GCATTTGGCT	GTATATAAGA	TACAAATGTG	TAAAAGTGTT	1320
AAAGCTCCAG	GAATATGTAT	TGATGTTGAT	AATGAAGATT	TGTTCTTTAT	AGCTGATAAA	1380
AATAGTTTTT	CAGATGATTT	ATCTAAAAAC	GAAAGAATAG	AATATAATAC	ACAGAGTAAT	1440
TATATAGAAA	ATGACTTCCC	TATAAATGAA	TTAATTTTAG	ATACTGATTT	AATAAGTAAA	1500
ATAGAATTAC	CAAGTGAAAA	TACAGAATCA	CTTACTGATT	TTAATGTAGA	TGTTCCAGTA	1560
TATGAAAAAC	AACCCGCTAT	AAAAAAAATT	TTTACAGATG	AAAATACCAT	CTTTCAATAT	1620
TTATACTCTC	AGACATTTCC	TCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	ATAAAAATAA	AATTATTAAA	2040
ACAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTTTAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
TATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280
aatgagggta	TTAACCAAGC	TATAGATAAT	TTAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA	TGAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
<b>TCAATATATA</b>	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

#### (2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 2574 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double
  (D) TOPOLOGY: linear

#### (ii) MOLECULE TYPE: DNA (genomic)

#### (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ATGCCAGTTA	CAATAAATAA	TTTTAATTAT	AATGATCCTA	TTGATAATAA	TAATATTATT	60
ATGATGGAGC	CTCCATTTGC	GAGAGGTACG	GGGAGATATT	ATAAAGCTTT	TAAAATCACA	120
GATCGTATTT	GGATAATACC	GGAAAGATAT	ACTTTTGGAT	ATAAACCTGA	GGATTTTAAT	180
AAAAGTTCCG	GTATTTTTAA	TAGAGATGTT	TGTGAATATT	ATGATCCAGA	TTACTTAAAT	240

ACTARIGATA AAAAGAATAT ATTTTTACAA ACAATGATCA AGTTATTTAA TAGAATCAAA	300
TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA	360
GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA	420
TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA	480
TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT	540
TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA	600
AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT	660
TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTCATCC ACGTCCTCCA CGGTCTCTAC	720
GGTATCAAAG TAGACGACCT CCCGATCGTC CCGAACGAAA AAAAATTCTT CATGCAGAGC	780
ACCGACGCAA TCCAGGCAGA AGAACTCTAC ACCTTCGGTG GTCAGGACCC GAGCATCATC	840
ACCCCGAGCA CCGACAAAAG CATCTACGAC AAAGTCCTCC AGAACTTCCG TGGTATCGTC	900
GACCGTCTCA ACAAAGTCCT CGTCTGCATC AGCGACCCGA ACATCAACAT CAACATCTAC	960
AAAAACAAAT TCAAAGACAA ATACAAATTC GTCGAAGACA GCGAAGGTAA ATACAGCATC	1020
GACGTCGAGA GCTTCGACAA ACTCTACAAA AGCCTCATGT TCGGTTTCAC CGAAACCAAC	1080
ATCGCAGAAA ACTACAAAAT CAAAACCCGT GCAAGCTACT TCAGCGACAG CCTCCCGCCG	1140
GTCAAAATCA AAAACCTCCT CGACAACGAA ATCTACACCA TCGAAGAAGG TTTCAACATC	1200
AGCGACAAAG ACATGGAAAA AGAATACCGT GGTCAGAACA AAGCAATCAA CAAACAAGCT	1260
TACGAAGAAA TCAGCAAAGA ACACCTCGCA GTCTACAAAA TCCAGATGTG CAAAAGCGTC	1320
AAAGCACCGG GTATCTGCAT CGACGTTGAC AACGAAGACC TCTTCTTCAT CGCAGACAAA	1380
AACAGCTTCA GCGACGACCT CAGCAAAAAC GAACGTATCG AATACAACAC CCAGAGCAAC	1440
TACATCGAAA ACGACTTCCC GATCAACGAA CTCATCCTCG ACACCGACCT CATCAGCAAA	1500
ATCGAACTCC CGAGCGAAAA CACCGAAAGC CTCACCGACT TCAACGTTGA CGTCCCGGTC	1560
TACGAAAAAC AGCCGGCAAT CAAAAAAATC TTCACCGACG AAAACACCAT CTTCCAGTAC	1620
CTCTACAGCC AGACCTTCCC GCTAGATATA AGAGATATAA GTTTAACATC TTCATTTGAT	1680
GATGCATTAT TATTTTCTAA CAAAGTTTAT TCATTTTTTT CTATGGATTA TATTAAAACT	1740
GCTAATAAAG TGGTAGAAGC AGGATTATTT GCAGGTTGGG TGAAACAGAT AGTAAATGAT	1800
TTTGTAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT	1860
GTTCCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTTGAA	1920
AATGCTTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTTAATA	1980
CCTGTAGTTG GAGCCTTTTT ATTAGAATCA TATATTGACA ATAAAAATAA AATTATTAAA	2040
ACAATAGATA ATGCTTTAAC TAAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTTATA CAATAAAAGA GGGAATGTAT	2160
AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2220
TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTTTA ATGATATAAA TTCTAAACTT	2280

WO 98/07864 PCT/GB97/02273

#### - 114 -

AATGAGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	234
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	240
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	246
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
<b>ICAATATATA</b>	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

#### **CLAIMS**

- 1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
- 2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
- 3. A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light chain.
- 4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
- 5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain  $H_{\rm N}$  portion.
- 7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain  $H_N$  portion.
- 8. A polypeptid according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

- 9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
- 10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
- 11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain  $\beta$  of Staphylococcal protein A.
- 12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
- 13. A polypeptide according to Claim 12 wherein said third domain is insulin-like growth factor-1 (IGF-1).
- 14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated  $H_{\text{N}}$  of a botulinum toxin heavy chain.
- 15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
- 16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

- 17. A polyp ptide according to Claim 14 wherein on or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
- 19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
- 21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
- 24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H<sub>c</sub> of a botulinum toxin h avy chain.

- 26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
- 27. A polypeptide according to Claim 26 lacking an intact portion designated  $H_{\rm c}$  of a botulinum toxin heavy chain.
- 28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
- 30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain  $H_N$  portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain  $H_N$  portion.
- 31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
- 32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.
- 33. A fusion protein according to Claim 32 wherein said s cond polypeptide is

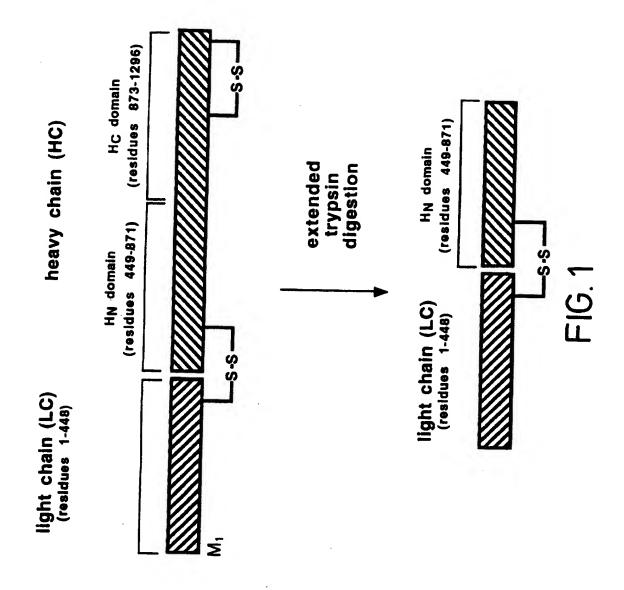
adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

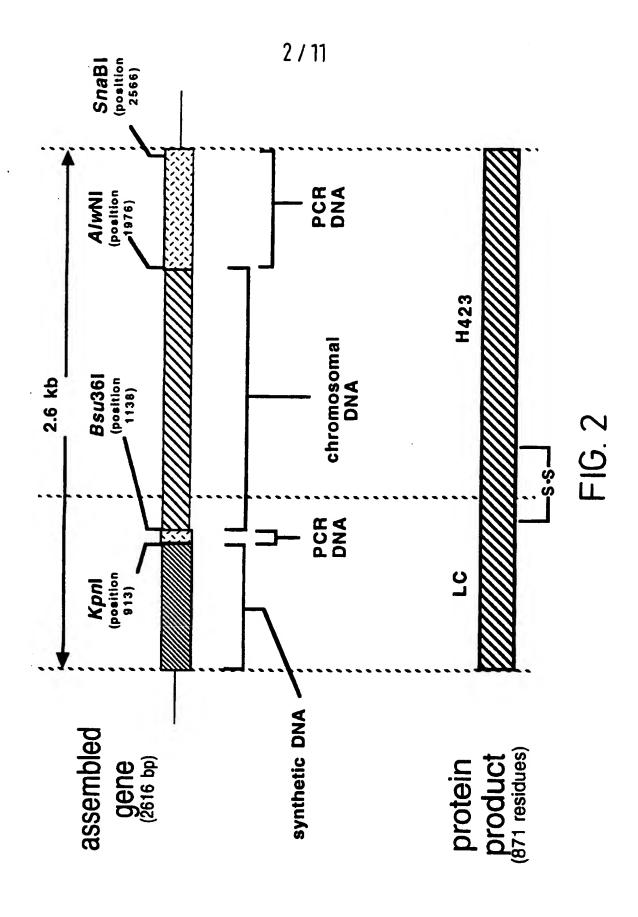
- 34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
- 35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.
- 36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.
- 37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.
- 38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.
- 39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.
- 41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim

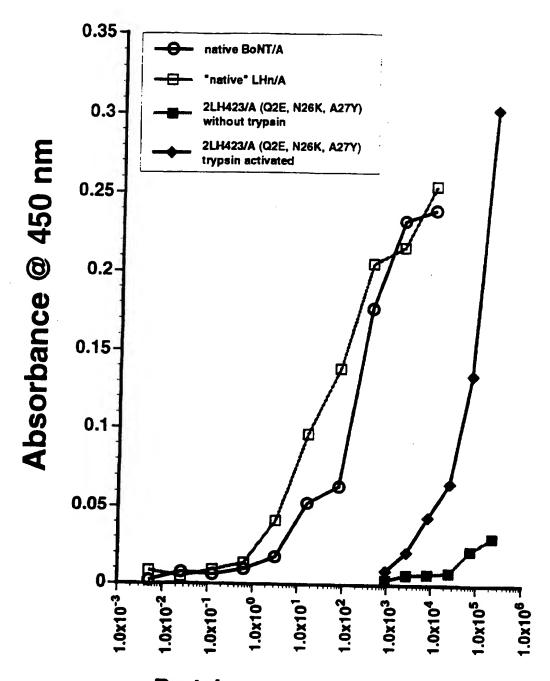
- 40 and comprising nucl otides encoding residues 1-448 of a botulinum toxin type A light chain.
- 42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain  $H_N$  domain.
- 43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
- 44. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain  $H_N$  domain.
- 45. A nucleic acid according to any of Claims 40-44 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
- 46. A nucleotide according to Claim 45 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.
- 47. A DNA according to any of claims 40-46.
- 48. A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24.
- 49. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.
- 50. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid encoding a fusion protein

according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

- 51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
- 52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.





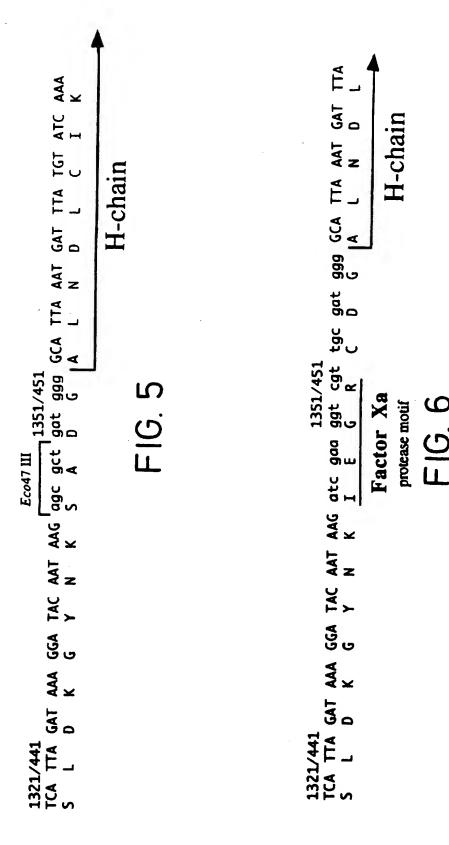


Protein concentration (ng/ml)

FIG. 3

DIAYIKIPNAGQMQPV (Seq 1.D. 2)		G Q M Q P V (Seq I.D. 4)	6 0 M Q P V (Seq I.D. 6)	> 4 0 M 0 5	6 Q M Q P V	
Z Z		א ר א	<b>≻</b>	۷ 2	۷ z	
			_	_	<u> </u>	
H		IKIP	X	н	¥	
<b>×</b>		×	¥	×	¥	
ı			H	H	H	
<b>4</b>		<b>≻</b>	<b>≻</b>	<b>≻</b>	<b>≻</b>	
H	۵	н	H	H	ı,	
<b>a</b>		۵	۵	9	0	
<b>&gt;</b> 9	ш —	>	>	>	>	
9	u.	5	G	G	9	
Z	ш	Z	Z	Z	Z	
> a	0 K L L	>	>	>	>	
0	<b>-</b>	<b>D</b>	0 P	0 P	0 P	
×	0	×	~	*	<u>~</u>	
<b>&gt;</b>	_	<b>&gt;</b>	<b>&gt;</b>	<b>&gt;</b>	<b>&gt;</b>	
Z	œ	Z	2	Z	Z	
LL	<b>-</b>	<u>.                                    </u>	щ.	ı.	ı.	
<b>0</b>	S	<b>0</b>	0	0	<b>⊘</b>	
<u> </u>	<b>5</b>	<b>*</b>	* <b>*</b>	<b>×</b>	~	
_ >	I	>	>	>	<b>-</b>	
<u>u</u>	H	ш	ш.	Ľ.	ıL	
MQFVNKQF 1	G	ш	ш	M Q F V N K Q F	MPFVNKQF	
<b>I</b>	GSPGIHMTST	GTMEFVNKQF - 24	GSMEFVNKQF 123	ΣH	Σ .⊣	
	ν	<b>—</b>	G S M 123			
			9 7			
LH <sub>423</sub> /A	23LH423/A	(Q <sub>2</sub> E, N <sub>26</sub> K, A <sub>27</sub> Y)	2LH423/A (Q2E, N26K, A27Y)	Native BoNT/A, C. botulinum 2169 Thompson et al. 1990	Native BoNT/A, C. <i>botulinum</i> 62A Binz <i>et al</i> .1990	

= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.



6/11

ACA T CAG GAC D TAT Y 66A 6 ATG M CCT P 6CG A GAG E 767 C GTG V AGG R CTG L AGT CGG S R AGG R TTC F CAG AGG R AGC S 6TG V AGC S TAT ر 30 GAG E F F TCA S ACA T GCT A <u>در</u> در AAG K 767 C 999 AAG K 767 **≸**~ 76C C AAC N GAG F CCT P CTC T F AAG K GAT D TAT Y GTG V 2587/863 TAC GTA Y V 2647/883 CCG GAG P E 2707/903 GGC TTT G F 2767/923 GGT ATC G I 2827/943 GCA CCC

=16.7

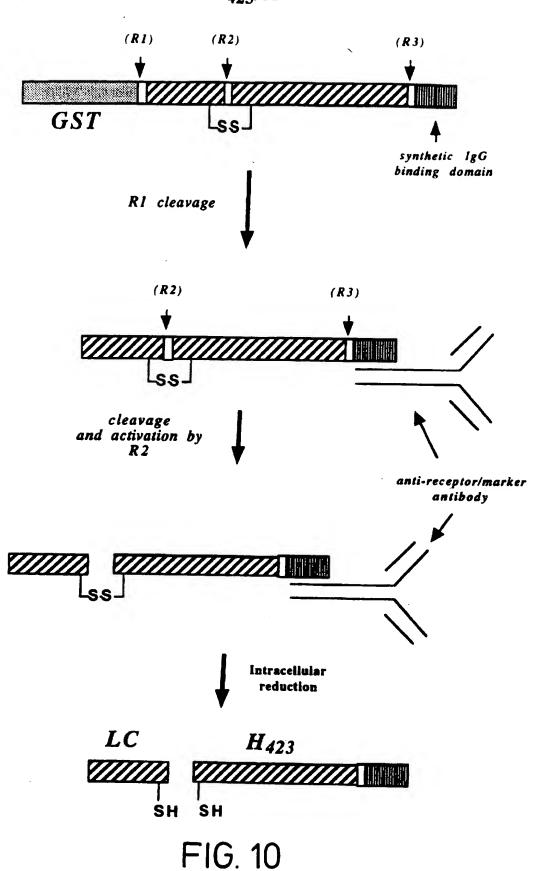
H CAT ACA T TCT S AAG K GAT D ATT I ATT I TAT GAT D GAA E 2617/873 TTT ACT F T 2677/893 TAT CAA Y Q ر و و ACA T TCT S S S Ħ ۲ ۲ ATA I ¥ L AGA R **₹**~ AGA R **₹**~ E GA A A A A 2587/863 TAC GTA GAT Y V D 2647/883 TCT AAA GTT S K V S AAG GAT K D

F16.8

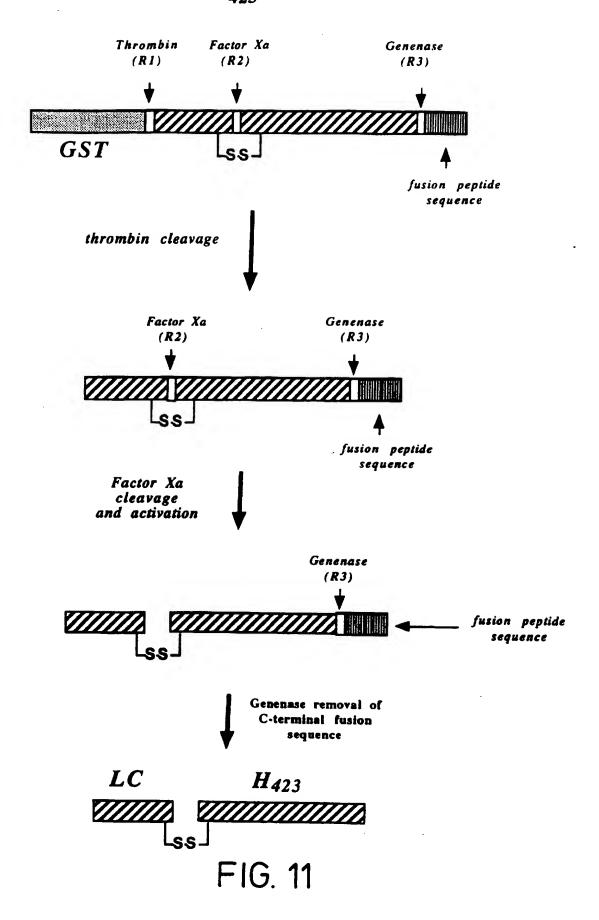
CAA CCA GCC AAC N A A GAA E GAA E TTA L AAA K C SA GAA E AAA X GAA E CAA O AAC N E\_ AAG K AAC N AAC GAA E GCT A A A A GAA E TAT Y AAC N AGC S AAC N AAC N AAC N 2617/873
TTT ACT GAA
F T E
2677/893
GAT GAA GCC
D E A
2737/913
CAT TTA CCT
H L P
2797/933
CCA AGC CAA
P S Q
2857/953
AAA GTA GAC
K V D A H GAC D GCG A GAT D ¥ F TTA L 900 A GAG F AAA 0 0 0 TTA L GAC D T L TAT Y ا≺ ¥ GCT A GAT CAA AGT ĘŦ GAT GAG E ₽ SCI **§** ~ 9 8 8 AAT N TAT Y ¥ z ATC I CTA TAC GTA GAT
Y V D
2647/883
TCC CCG GGT
S P G
2707/903
GAA CAA CAA
E Q Q
2767/923
AAC GCC TTC
N A F
2827/943
GCT AAA AAG
A K K
2887/963
CAA AAC GCG
Q N A
2947/983
TTC ATC CAA
F I Q
3007/1003
AAG CTA AAT
K I N

F16.9

8/11 LH<sub>423</sub>/A



# $LH_{423}/A^{9/11}$



10/11

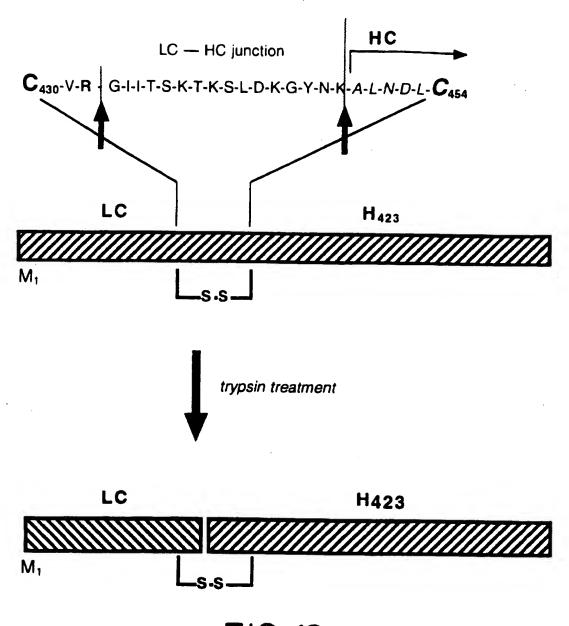
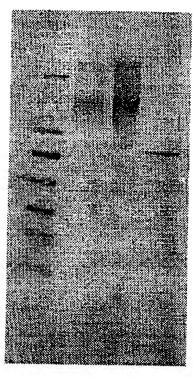


FIG. 12

# 11 / 11

Panel A. 1 2 3 4



Panel B.

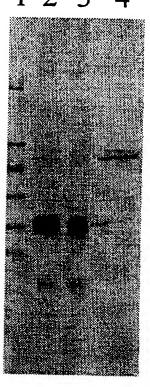


FIG. 13

#### INTERNATIONAL SEARCH REPORT

Inten Jnal Application No PCT/GB 97/02273

PCT/GB 97/02273 CLASSIFICATION OF SUBJECT MATTER PC 6 C12N15/31 C12N IPC 6 C12N1/21 C12P21/02 C07K14/33 A61K38/16 A61K39/08 According to International Patent Classification (IPC) or to both national classification and IPC B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C12P A61K Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No. X WO 96 12802 A (OPHIDIAN PHARM INC 1-52 ;WILLIAMS JAMES A (US); PADHYE NISHA V (US); KI) 2 May 1996 see the whole document X KURAZONO H ET AL: "Minimal essential 1-52 \*domains\* specifying toxicity of the \*light\* \*chains\* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910 see table II -/--Further documents are listed in the continuation of box C. Χl Patent family members are listed in annex. Special categories of cited documents : \*T\* later document published after the international filing date or priority date and not in conflict with the application but "A" document defining the general state of the art which is not considered to be of particular relevance cited to understand the principle or theory underlying the "E" earlier document but published on or after the international "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered. filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu-"O" document referring to an oral disclosure, use, exhibition or other means ments, such combination being obvious to a person skilled \*P\* document published prior to the international filing date but later than the priority date claimed "&" document member of the same patent family Date of the actual completion of the international search Date of mailing of the international search report 3 0. or 98 9 December 1997 Name and mailing address of the ISA **Authorized officer** European Patent Office, P.B. 5818 Patentiaan 2 NL - 2280 HV Rijewijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni, Fax: (+31-70) 340-3016 Hillenbrand, G

### INTERNATIONAL SEARCH REPORT

Interconal Application No
PCT/GB 97/02273

		PCT/GB 97/022/3						
C.(Continue	(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT							
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.						
X	LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document	1						
A -	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document	1,26,35						
		·						

## INTERNATIONAL SEARCH REPORT

PCT/GB 97/02273

Patent document cited in search report	Publication date	Patent family	Publication
	uale	member(s)	date
WO 9612802 A	02-05-96	US 5601823 /	A 11-02-97
		US 5196193 /	A 23-03-93
		AU 3968395 /	A 15-05-96
		EP 0796326 /	A 24-09-97
		FI 971732 /	A 23-06-97
		NO 971868 A	
		PL 320214 A	
		ZA 9508990 A	
		AU 6653894 A	A 04-07-94
		CA 2150935 A	23-06-94
		EP 0671902 A	
		WO 9413264 A	
		US 5466672 A	
		US 5599539 A	
		AU 638786 B	
		AU 6895191 A	
		EP 0498854 A	
		WO 9106306 A	
		US 5443976 A	
		US 5340923 A	